

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:26:27 ; Search time 50 Seconds
(without alignments)

8773.789 Million cell updates/sec

Title: US-09-126-816b-6_COPY_1_1700

Perfect score: 8677

Sequence: 1 MNLVYKALQKQWVYKFRIQ.....YLYGIDRYVNVKVIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8646	99.6	2364	2 Q46342	Q46342 clostridium
2	6761	77.9	2367	2 Q9EXR0	Q9EXR0 clostridium
3	6732	77.6	2367	2 Q9FP31	Q9FP31 clostridium
4	6711	77.3	2367	2 Q46034	Q46034 clostridium
5	2144	24.7	554	2 Q93L39	Q93L39 clostridium
6	2070.5	23.9	2178	2 Q46149	Q46149 clostridium
7	1911	22.0	698	2 Q68653	Q68653 clostridium
8	1893	21.8	698	2 Q9EXQ8	Q9EXQ8 clostridium
9	1887	21.7	697	2 Q86141	Q86141 clostridium
10	1469	16.9	553	2 Q9FCX5	Q9FCX5 clostridium
11	501.5	5.8	3169	2 Q9ZGR4	Q9ZGR4 escherichia
12	497	5.7	2792	5 Q814R2	Q814R2 plasmodium
13	496.5	5.7	3169	2 Q82916	Q82916 escherichia
14	494	5.7	3317	16 Q8EWP8	Q8EWP8 mycoplasma
15	478	5.5	2771	5 Q26216	Q26216 plasmodium
16	463	5.3	2752	5 Q9BJY0	Q9BJY0 plasmodium

17	460	5.3	5767	5 Q81525	Q81525 plasmodium
18	446	5.1	3130	5 Q9BK46	Q9BK46 plasmodium
19	446	5.1	3130	5 Q81DX6	Q81DX6 plasmodium
20	446	5.1	3254	5 Q9BK45	Q9BK45 plasmodium
21	439.5	5.1	2747	5 Q9BJX9	Q9BJX9 plasmodium
22	437	5.0	2269	5 Q26223	Q26223 plasmodium
23	436.5	5.0	2849	5 Q81HY4	Q81HY4 plasmodium
24	436	5.0	3381	5 Q812V4	Q812V4 plasmodium
25	427	4.9	2184	5 Q81LO6	Q81LO6 plasmodium
26	424	4.9	2166	16 Q51465	Q51465 borrelia bu
27	423.5	4.9	4688	16 Q9PQ08	Q9PQ08 ureaplasma
28	410.5	4.7	6761	5 Q81C77	Q81C77 plasmodium
29	406.5	4.7	1711	5 Q8MW2	Q8MW2 plasmodium
30	406.5	4.7	1956	5 Q81IE1	Q81IE1 plasmodium
31	404.5	4.7	1713	5 Q8MW1	Q8MW1 plasmodium
32	404.5	4.7	3597	5 Q81LR5	Q81LR5 plasmodium
33	403	4.6	1686	5 Q81FP9	Q81FP9 plasmodium
34	403	4.6	3225	16 Q9PKM6	Q9PKM6 chlamydia m
35	399.5	4.6	2511	5 Q81L44	Q81L44 plasmodium
36	399.5	4.6	3322	5 Q81KLO	Q81KLO plasmodium
37	398	4.6	1979	5 Q96133	Q96133 plasmodium
38	397	4.6	5687	5 Q81JH4	Q81JH4 plasmodium
39	396.5	4.6	3223	2 Q81LL9	Q81LL9 escherichia
40	396	4.6	1697	5 Q81FM4	Q81FM4 plasmodium
41	396	4.6	1716	5 Q8MWH2	Q8MWH2 plasmodium
42	394.5	4.5	3223	2 Q9RPH1	Q9RPH1 escherichia
43	394.5	4.5	6473	5 Q81KH9	Q81KH9 plasmodium
44	394	4.5	2867	5 Q9N2M3	Q9N2M3 plasmodium
45	394	4.5	2910	5 Q81BY8	Q81BY8 plasmodium

ALIGNMENTS

RESULT 1

Q46342 ID Q46342 PRELIMINARY; PRT; 2364 AA.
 AC Q46342;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Cytotoxin L.
 OS Clostridium sordellii.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1505;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN=6018;
 RX MEDLINE=95369733; PubMed=7642137;
 RA Green G.A., Schue V., Montell H.;
 RT "Cloning and characterization of the cytotoxin L-encoding gene of
 RT Clostridium sordellii: homology with Clostridium difficile cytotoxin
 RT B.";
 RL Gene 161:57-61(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6018;
 RX MEDLINE=96149194; PubMed=8544213;
 RA Green G.A., Schue V., Girardot R., Montell H.;
 RT "Characterisation of an enterotoxin-negative, cytotoxin-positive
 RT strain of Clostridium sordellii.";
 RL J. Med. Microbiol. 44:60-64(1996).
 DR EMBL; X82638; CAA57959.1; -;
 DR InterPro; IPR002479; CW binding.
 DR InterPro; IPR001950; Trf_SUI1.
 DR Pfam; PF01473; CW binding_1; 18.
 DR PROSITE; PS01118; SUI1_1; 1.
 SQ SEQUENCE 2364 AA; 270576 MW; EAD8A4467A89BDBB CRC64;

Query Match 99.6%; Score 8646; DB 2; Length 2364;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1696; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNLVNAQLOKQVYKFRIOEDYVAAILNALBEYHNMSSESVVEKYKLKDINNTDNYL 60
DB 1 MNLVNAQLOKQVYKFRIOEDYVAAILNALBEYHNMSSESVVEKYKLKDINNTDNYL 60
QY 61 NTYKSGRNKALKKPKXYLTMEVLELKNNSLPVPEKNLHFIWIGQINDTAINYNQWKD 120
DB 61 NTYKSGRNKALKKPKXYLTMEVLELKNNSLPVPEKNLHFIWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVFYDSNAPLINTLTKTIVESATNNTLESFRENLDNDFDYNNKFRKMEIYY 180
DB 121 VNSDYTVKVFYDSNAPLINTLTKTIVESATNNTLESFRENLDNDFDYNNKFRKMEIYY 180
QY 181 DQKQHPIDYKSOIENPEFIIDNIKTYSNEYSKDLEALNKYIIESLKNKITANNNDI 240
DB 181 DQKQHPIDYKSOIENPEFIIDNIKTYSNEYSKDLEALNKYIIESLKNKITANNNDI 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIQDLFKS 300
DB 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIQDLFKS 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQSFESALSSKSKSEIF 360
DB 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQSFESALSSKSKSEIF 360
QY 361 LPDDDIKVSPLVKIAFANNSVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE 420
DB 361 LPDDDIKVSPLVKIAFANNSVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKIFSDKLASISNEDNMFMKIITNYLKVGFPADVRSTINLSGPGVYTGAYOD 480
DB 421 GTDFNTMTKIFSDKLASISNEDNMFMKIITNYLKVGFPADVRSTINLSGPGVYTGAYOD 480
QY 481 LMFKXDNSTNIHLEPELNFPPKTKISOLTEQETISLWSFNQARAKSQFEYKKGYPF 540
DB 481 LMFKXDNSTNIHLEPELNFPPKTKISOLTEQETISLWSFNQARAKSQFEYKKGYPF 540
QY 541 GALGEDNDLPQNTVLDKDYYSKILSSMKTRNKEYIHYIVLOQDKISYEASCNLFSK 600
DB 541 GALGEDNDLPQNTVLDKDYYSKILSSMKTRNKEYIHYIVLOQDKISYEASCNLFSK 600
QY 601 DPYSSILYQKNEGSETAYYYVADAEIKEIDKIRIPYQISNKRNIKLTFFIGHGSEFNT 660
DB 601 DPYSSILYQKNEGSETAYYYVADAEIKEIDKIRIPYQISNKRNIKLTFFIGHGSEFNT 660
QY 661 DTFANLDVDSLSSEIETIILNLAADISPKYIEINLLGCNMFYSIYAEITYPGKULLKIK 720
DB 661 DTFANLDVDSLSSEIETIILNLAADISPKYIEINLLGCNMFYSIYAEITYPGKULLKIK 720
QY 721 DRVSELMPSISODSITVSANQYEVRIENERGKEIILDHSGKWTNKEESIIDKISSKEYISF 780
DB 721 DRVSELMPSISODSITVSANQYEVRIENERGKEIILDHSGKWTNKEESIIDKISSKEYISF 780
QY 781 NPKNENKIIIVKSKYLHELSTLLOEIRNANSDDIDLEKKVMLTECEINVASNIDRQIVEGR 840
DB 781 NPKNENKIIIVKSKYLHELSTLLOEIRNANSDDIDLEKKVMLTECEINVASNIDRQIVEGR 840
QY 841 IBEAKNLTSDSINYIKNEPKLIESDSLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
DB 841 IBEAKNLTSDSINYIKNEPKLIESDSLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
QY 901 INKETGNSIFETEKEIPSEYATHISKEISNKTOTIFDNVNGKLVKKNLDAAEVNTLN 960
DB 901 INKETGNSIFETEKEIPSEYATHISKEISNKTOTIFDNVNGKLVKKNLDAAEVNTLN 960
QY 961 SAFFIQSLLEYNTTKESLSNLSVAMKVQVYVAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 SAFFIQSLLEYNTTKESLSNLSVAMKVQVYVAQLFSTGLNTITDASKVVELVSTALDETID 1020
QY 1021 LPLTLEGLPIIATIIDGVSGLGAAIKELSETNDPLLRQIEIAKIGIMAVNLTAATAIYV 1080
DB 1021 LPLTLEGLPIIATIIDGVSGLGAAIKELSETNDPLLRQIEIAKIGIMAVNLTAATAIYV 1080

QY 1081 SALGIASGSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGAFITLD 1140
DB 1081 SALGIASGSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGAFITLD 1140
QY 1141 DKIIIMPQDDLVLSEIDFNNSITLGKCEIMRAEGSGHTLTDDIDHFFSSPSITYRKPW 1200
DB 1141 DKIIIMPQDDLVLSEIDFNNSITLGKCEIMRAEGSGHTLTDDIDHFFSSPSITYRKPW 1200
QY 1201 SIYDVLANIKKEDIDFSDKDLMLPNAPRVFGYEWGTFGFRSLNDGDKLDRIDHYEG 1260
DB 1201 SIYDVLANIKKEDIDFSDKDLMLPNAPRVFGYEWGTFGFRSLNDGDKLDRIDHYEG 1260
QY 1261 QFYWRYPAFIADALITKLKPRYEDTNVRLNDGNTRSPFIVPITTEQIRKXLSYFYSG 1320
DB 1261 QFYWRYPAFIADALITKLKPRYEDTNVRLNDGNTRSPFIVPITTEQIRKXLSYFYSG 1320
QY 1321 GSYLSLSGSPYNNIDNLVENDTWIDVNVVKNITTESDEIQKGLIENILSKLNIEDN 1380
DB 1321 GSYLSLSGSPYNNIDNLVENDTWIDVNVVKNITTESDEIQKGLIENILSKLNIEDN 1380
QY 1381 KIILNNHTINPYGDNESNRFISLTFSLEIDNIIIEIDVSKSYKILLSGNCKLIENS 1440
DB 1381 KIILNNHTINPYGDNESNRFISLTFSLEIDNIIIEIDVSKSYKILLSGNCKLIENS 1440
QY 1441 SDIOQKIDHIGFNGEHQKYIFVSYIDNETKNGFIDYSKKEGLFTAEPFNSIESIIRNIYMP 1500
DB 1441 SDIOQKIDHIGFNGEHQKYIFVSYIDNETKNGFIDYSKKEGLFTAEPFNSIESIIRNIYMP 1500
QY 1501 DSNLLFYSSKDLKDRIIRINKGDKVLLIGNFYKDDMKVLSFTTIEDTNTIKLNGVYLDEN 1560
DB 1501 DSNLLFYSSKDLKDRIIRINKGDKVLLIGNFYKDDMKVLSFTTIEDTNTIKLNGVYLDEN 1560
QY 1561 GVAQILKFMNNAKSALNTSNLMFLESINIKNIPYNNLDNPIERFILTNTFIISGNSIG 1620
DB 1561 GVAQILKFMNNAKSALNTSNLMFLESINIKNIPYNNLDNPIERFILTNTFIISGNSIG 1620
QY 1621 QFELICDKDKNIQPIYFNFKIKETSYTLVGNRQNLIVERPSYHLDDSGNISSTVINFSQK 1680
DB 1621 QFELICDKDKNIQPIYFNFKIKETSYTLVGNRQNLIVERPSYHLDDSGNISSTVINFSQK 1680
QY 1681 YLYGIDRYVNVKVIITAPNLYT 1700
DB 1681 YLYGIDRYVNVKVIITAPNLYT 1700
RESULT 2
Q9EXR0 PRELIMINARY; PRT; 2367 AA.
AC Q9EXR0;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Toxin B.
GN TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RC STRAIN=8864;
RP SEQUENCE FROM N.A.
RA von Eichel-Streiber C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA Kohl M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011301; CAC19891.1; -;
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR001950; TIF SU11.
DR Pfam; PF01473; CW_binding_1; 17.

DR	PROSITE; PS01118; SUIL_1; 1.
SQ	SEQUENCE 2367 AA; 269072 MW; BF7087C92E7388E7 CRC64;
	Query Match 77.9%; Score 6761; DB 2; Length 2367;
	Best Local Similarity 76.4%; Pred No 5.6e-255;
	Matches 1300; Conservative 200; Mismatches 199; Indels 2; Gaps 2;
Qy	1 MNLVNAQLQRMVVYKFRIDEYVAIVLALAEYYHNHNSSESVBEKYLKJLDINNLDTNYL 60
Db	1 MSLVNRKQLEKANVRPRVQEDEVAILDALAEYYHNHNSNTVVEKYLLKDINSLTDTYI 60
Qy	61 NTYKKSGRNKALKFKPKLYLTWEVLSEKONSITPVERKLHFIIWGGQINDTAINYINQWKD 120
Db	61 DTYKKSGRNKALKFKPKLYLTETELSEKNSLTPVERKNLHFIIWGGQINDTAINYINQWKD 120
Qy	121 VNSDYTVKFVVDNSAFLINTLKTTIVESATNNTLESFRENLNDDPEPDYNKFFRYKRMEIY 180
Db	121 VNSDYNVNVDNSAFLINTLKTTIIESANDTLESFRENLNDDPEFHRTAFRRKMQLIY 180
Qy	181 DKQKHFDIYKQSIEENPEFIIDNIITYLSNEYSKOLEALNKIYEESLANKITANNNGDI 240
Db	181 DKQNFINYYKAQKEENPDLIIDDIVKTYLSNEYSKOIDELNAVIEESLNKVTSNGDV 240
Qy	241 RNLEKFADEDLRLYNQELVERWNLAASDIRLSMLKEDGCVYLDVDIILPGIDPLFKS 300
Db	241 RNPBFPKTGEVFNLYEQELVERWNLAGASDIRVAIKNIGGVYLDVDMLPGIHPDLPKD 300
Qy	301 INKPDSI-TNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVORSFESALSCKSEI 359
Db	301 INKPDSVKTAVDWEQMQLAIMKYKEYIPETSXHPFTLDEEVOSFESVLASCKSEI 360
Qy	360 FLPLDDDKVSPLEVAKIFANNSVINQALISLKDSYCSDLVINQIKRYKIILNDLNPSIN 419
Db	361 FLPLGGIEVPLEVKVAFAGSIIDQALISAKDSYCSDDLAIKQIQNRYKIILNDTLGLPIIS 420
Qy	420 EGDTFNTMKIPSDKLASINEDNMFMWIKITNLYKVGAPADVSTTNLSGPVGVTGAYQ 479
Db	421 QGDNFTNNFGESLGAIAENEINSPIAKIGSYLRVGFYPEANTTTLSGPTTIYAGAYK 480
Qy	480 DLLMFKDNSNTIHILEPEARNFEPKTKISQLTEQBITSLSWSNQARAKSQFEYKKGYP 539
Db	481 DLITFKMSIDTSLSSSELNRNFPKPNVISQATEQRKNSLMQFNEERAKIQFEYKKNYP 540
Qy	540 EGALGEDDNLDFAQNTVLDKDYVSKKILSSMKTNRKEYIHYIVLOQDGKISYEASCNLF 599
Db	541 EGALGEDDNLDFSQNTVTDXLEYLLEKISSTSKSSRGVHYIVVOLQGDKISYEACNLPA 600
Qy	600 KDPVSSILYQKNIGSETATYYYYVADAIEIKEDIRPYQISNKNRIKLTFIGHGKSEFN 659
Db	601 KNPYDSILFQKNIEDSEVAYYNPTDSEIOBKDYRIPDRISDRPKIKLTLIGHGKAEFN 660
Qy	660 TDPANILDVLSSEIETILNLAKADISPKYIEINLLGCNMFYSYIAEETYPCGLALKI 719
Db	661 TDIPAGLDVLSSEIETILDOLAKADISPKSIEINLLGCNMFYSVNVEETYPCKLLRV 720
Qy	720 KDRVSELMPSISODSITVSANQYEVREINEEGKREILDHSKWINKBSIINDISSKEYIS 779
Db	721 KDRVSELMPSISODSIIVSANQYEVREINSEGRRELLDHSGEWINKESSIIOKISSKEYIS 780
Qy	780 FNPKENKIIVKSKYLHELSTLLQEIERNANSDDILEKKVMLTCEENVASNIDROQVEG 839
Db	781 FNPKENKIIIVKSKNLPBLSTLLQERIINNSNSDIELEEKVWLACEENVIISNIETQVVEE 840
Qy	840 RIEAKNLTSDSNINYIKNEFKLIESIDSLSYDLKHONGLLDSSHIPSIEDISKTEGNFRIR 899
Db	841 RIEBAKSLTSDSNINYIKNEFKLIESISDALYDLQOQNELEBESHIPSIEDISKTEGFSIR 900
Qy	900 FINKETGNSPIETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAAEHVNTL 959
Db	901 FIDKETGESIFVTEKAI FSEYANHIIITEISKLDKDTIFDVNGKLVKKVNLDATEHVNTL 960
Qy	960 NSAFFIQSLIEYNTKTESLNSLVAMKVQVYAQLFSTGLCNTITTDASKVELVSTALDETI 1019

Db	961	NAAFFIQSLIGVYNSKESLSNLVAMKVQVYAOLFSTGLNTIITDAAKVVELVSTALDETTI	1021
Qy	1020	DLLEPTLSEGLPIITATIIDGVSGLAAIKELSETNDPLLRQIEIAKIGIMAVNLTAASTAIV	1079
Db	1021	DLLEPTLSEGLPVTATIIDGVSGLASIKELSETSDPLLRQIEIAKIGIMAVNLTAATAII	1080
Qy	1080	TSALGIASGPSILLVPLAGISAGIPSLVNNELILOKATKVIDYFKHISLAETEGAPTLL	1139
Db	1081	TSSLGIASGPSILLVPLAGISAGIPSLVNNELILRAEAKNVVDYFGHISLAESSEGAPTLL	1140
Qy	1140	DDKIIMPODDLVLSEIDFNNNSTLTKCEIWRAGGSGHTLTDIDHFFSSPSITVYRKPW	1199
Db	1141	DDKIIMPODDLVLSEIDFNNNSTLTKCEIWRMEGGSGHTVTDIDHFFSAPSTTVREPY	1200
Qy	1200	LSIYDVNLTKKEKIDFSSKOLMVLNPNRVPFGYEMGWTGCFRSLDNDGTKLLDRIDRHYE	1259
Db	1201	LSIYDVLDVKEBELDLSKOLMVLNPNADRIFGHERGWTFCRLSLENDGTKLLDRIDRHYE	1260
Qy	1260	GQYWRYPAPIADALITKUKPRVEDTVNRINLGNTRSFIVPVITTEQIRKNI.SYSFPGS	1319
Db	1261	GQYWRYPAPFADSVITKUKPRVEDTVNRISLDSNTRSFIVPVITTEYIREKLSYSFPGS	1320
Qy	1320	GGSYSLSPYNNIIDLNVENDTWIDVDNVVKNITIESDEIQKGLIENILSKLNIED	1379
Db	1321	GGTYALSLSQYNNINIELNENDTWIDVDNVVRDVTIESDKIKKGLIENILSKUSIED	1380
Qy	1380	NKTIILNHNTHFYGDINESNRPIFLSFILSDINIIIEIDLVSQYKILLSGNCMKLIEN	1439
Db	1381	NKIILDNHINFSGTLNGGNGFVSLFFSILEGINAVIEVDLLSKSYKVLISGELKTLMAN	1440
Qy	1440	SSDIQOKIDHIGPNGBHQKIFYYSIDNEFKYNGPIDYSKKEGLFTAFBNESIIRNIYM	1499
Db	1441	SNSVQOKIDYIGLNSLQKNI.PYSFMDDEKGKNGFNCFTKEGLFVSELSDVVLLIKVYM	1500
Qy	1500	PDSNLPIYSSKDLDIRIINKGDVKLLIGNYFKDDMKVYLSLTIEDTNTIKLNGVYLDE	1559
Db	1501	DNSKPPFGYYSNDLKDVKTKDDVITITGYLLKDDIKISLFTIQDKNTIKLNGVYLDE	1560
Qy	1560	NGVAQILKPMNNAKSALNTSNLNMFLSINIKNI.FVNNLDPNIEFTLDTNFIISGNSNI	1619
Db	1561	NGVAELIKPMNK-KGSTNTSDLSMSFLESNNIKSIFIKLSKNAKILLDTNFIISGTTFI	1619
Qy	1620	GQFELICDKDKNIQPIYFNPKIKETSYTLVGNRQNLIVEPSYHLLDSDGNISSTVINFSQ	1679
Db	1620	GQFELICDKDKNIQPIYFNPKIKETSYTLVGNRQNLIVEPNTYHLLDSDGNISSTVINFSQ	1679
Qy	1680	KYLYGIDRYVNVKVIILAPNLYT	1700
Db	1680	KYLYGIDSCVNVKVIISPGIYT	1700
RESULT 3			
Q9F931 PRELIMINARY; PRT; 2367 AA.			
ID	Q9F931	PRELIMINARY; PRT; 2367 AA.	
AC	Q9F931;		
DT	01-MAR-2001	{T-EMBLrel. 16, Created}	
DT	01-MAR-2001	{T-EMBLrel. 16, Last sequence update}	
DT	01-MAR-2003	{T-EMBLrel. 23, Last annotation update}	
DE	Cytotoxin B.		
GN	TCDB.		
OS	Clostridium difficile.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1496;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=5340;		
RC	MEDLINE=20448897; PubMed=10924443;		
RA	Sambol S.P., Merrigan M.M., Lysterly D., Gerdling D.N., Johnson S.;		
RT	"Toxin gene analysis of a variant strain of clostridium difficile that		
RL	causes human clinical disease.";		
RL	Infect. Immun. 68:5480-5487(2000).		
DR	EMBL; AF217292; AAG18011.1; -.		

Qy	960	NSAFFQTOSLIETYNTTKESLNLSVAMKVQYIAQLFSTGLNLTITDASKVVELSVTALDETI	1019
Dd	961	NAAFFIQSLIETYSNESKESLNLSVAMKVQYIAQLFSTGLNLTITDAARAVVELSVTALDETI	1020
Qy	1020	DLPLPTISEGPIIATIIDGVSLGAALKELSETNDPLLRLRQBIEAKIGIMAVNLTAASAIIV	1079
Dd	1021	DLPLPTISEGPIIATIIDGVSLGAALKELSETSDPLLRLRQBIEAKIGIMAVNLTTATAII	1080
Qy	1080	TSALGIASGSFIILLVPLAGISAGIPSLVNNELIQDKATKVIDIDFKHISLAETEGFTLL	1139
Dd	1081	TSSLGIASGSFIILLVPLAGISAGIPSLVNNEVLVRDKATKVVDYFKHVSLETVEGVFTLL	1140
Qy	1140	DDKIIMPQDLVLSEIDTFNNNSTILGKECEIWRABGGSGHTLTDODIHFFSPSPITVKPW	1199
Dd	1141	DDKVMPPQDDLVLSEIDTFNNNSVLGKECEIWRMEGGSGHTVTDDIDHFFSAPSITTYREPH	1200
Qy	1200	LSIYDVNLINKKEKIDFSKOLMWLPNAPNRVFGEWMTGCFRSILDNGTKLLDRDRHYE	1259
Dd	1201	LSIYDVLEVQEELDUSKOLMWLPNAPNRVFAWETGTGRLSENDGTKLLDRIRNYE	1260
Qy	1260	GQFYWRYPAFIADALITKLPRYEDTNVRINLDGNSTRSFIVPVITTEQIRKNLSYSFGS	1319
Dd	1261	GEFWRYFAFIADALITTLKPRYEDTNIRINLDSNTRSFIVPITTEYIREKLSYSFGS	1320
Qy	1320	GGSYSLSLSPYNNIDLNVENDTWIDVDNVVNKITIESDETOKGELIENILSKLNIED	1379
Dd	1321	GGTYALSLSQYNMGINIELSESDDVIIDVDNVVDVTIESDKIKKGDLIEGLISTLSIEE	1380
Qy	1380	NKIILNHTNFYGDINDINESNRIFSLFSLIEDINIIEIDLVSKEYKILLSGNCMKLIEN	1439
Dd	1381	NKIILNSHEINFSEYVNGSNGFVSLFSLILEGINAIEVDLLSKSYKLLISGELKILMLN	1440
Qy	1440	SSDIQQIKDHIGFNGBEHQKIFYYSIDNETKYNGPIDYSKEGELFTAERFNSESIIRNYM	1499
Dd	1441	SNHIQQIDYIGFNSELQKNIPIYSFDVSEGKENGFIINGSTKEGFLVSELPDVVLLISKVYM	1500
Qy	1500	PDSNNLFIYSKKDKDIRIINKGDVKLLIGNYFKDDMKVSLSTIEDNTWIKLVGVYLDE	1559
Dd	1501	DDSKPSFGYYSNLKDVKVITKONVNILTGYILKDDIKISLUTLQDEKTIKUNSVHLDE	1560
Qy	1560	NGVAQILKFMMNAKSALTNSLMNFLESINIKNI FYNNLDPNIEFILDTNFIISGSNSI	1619
Dd	1561	SGVAELLKPMNR-KGSTNTSDLSMFLESNNIKSI FWNFLQSNIKFI LDANFII SGTTSI	1619
Qy	1620	GQFELICDKDNKIOPYFINPKIKETS YTLVGNRNQLIPEPSYHLDSDGNSISSTVINFSQ	1679
Dd	1620	GQFEFICDENNNIQPYIEKFTLENTLYTVGNRQNMI VEPNYDLDSDGISSTVINFSQ	1679
Qy	1680	KYLGGIDRYNKVIIAENLYT 1700	
Dd	1680	KYLGGIDSCVKNKVVISPNYT 1700	
RESULT 4			
ID	Q46034	PRELIMINARY;	PRT; 2367 AA.
AC	Q46034,		
DT	01-NOV-1996	(TrEMBLrel. 01, Created)	
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)	
DE	Toxin B.		
OS	Clostridium difficile.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
CC	Clostridium.		
NCBI_TaxID=1496;			
[1]			
RN	SEQUENCE FROM N.A.		
RC	STRAIN-isolate 1470:		
RX	MEDLINE=96079281; PubMed=7494480;		
RA	von Eichel-Streiber C., Meyer zu Heringdorf D., Habermann E.,		
RA	Sartinghen S.;		
RT	"Closing in on the toxic domain through analysis of a variant		

Db 763 DYFGKWSNTDLIAEQISKYKVVYVWNEVENTLSARVEQLNKVAEPAKDI-----NSIIQT 817
Qy 816 EKKVMTCEINVAAGNDQIVGEBRIEAKNLTSDSI---NYIKNEFKLTESIDSLYD 871
Db 818 TNNQELKSLVNTYADLTITLYSELKEDDIPFELDNIOIKERILNEISRHDPSNIID 877
Qy 872 LKHQGLDSDHIFSPEDISKTEGNSFRIRPINKETGNSIFIEKEIFSEYATHISKEISN 931
Db 878 FYQKNISNNMIILPDSIIKEKDYVNVKLANKITGETSVIKTSYDSLWNFTNKYKIVDD 937
Qy 932 IKDTITFDNVNKLKVKVNLDAAEVNTLSAPFIOSLIEYNTTKESLNSLVAMKVQVYA 991
Db 938 IKGIIVDQINGEFIKKAEFEQNPSSLNSAMQLLDIDYKPYTEBILTNMNTSLKVQYA 997
Qy 992 QLFSTGLMTITDASKVVELVSTALDETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSET 1051
Db 998 QIFQLSIGAIQATEIIVIIISDALNANFNILSKLVGSSVASVIIIDGINLAAALTELKVV 1057
Qy 1052 NDPLLRQIEAKIGIMAVNLTAASAVITSALG---IASGFSILVPLAGISAGIPSLVN 1108
Db 1058 KTNFERKLEIAKVMYSIGFLESLSLGLGATAVSEILGVISVPVAGILVGLPSLVN 1117
Qy 1109 NELIQQDKATKVIDYFKHISLAETEGATLDDKLIIMPQDDLVLSEIDPNNSITLGKE 1168
Db 1118 NILVGEKYNQILDYFSKPYFIVGNKPPFS-IQDNIIIPYDDIAITELNFKYKGYAK 1176
Qy 1169 IWRAGEGSGHTLTDIDHFFSPSTYTRKPMLSIYDVNLNKKIKEDFSKDLMLWLPNAPNR 1228
Db 1177 ISGLKVLGVLTHIGENIDHYFAPSUDHYIE-LSIYPALKLNDTNLP-KGNVVLLPSGLNK 1234
Qy 1229 VFGYEMGWTPGPRSLDNDGTLKLLDIRDHY-----EGQFYWRYFAFIADALITKLKPYE 1283
Db 1235 VYKPEISALAGANSQEGGVEVLNIRNYVDSNGNTKPPKYEAPF-EYSFYSRVEYF 1293
Qy 1284 DTVNRINDGNTRSFIVPITTEQIRKNLSYSFYGGSGYSLSUSPYNNIDNLNVENDT 1343
Db 1294 DTKVNVILDENKTLIIPLVATIDERNKISYEILGDGQYVILPVNQTINIVSNKNDI 1353
Qy 1344 WVIDVNVKNITTESDEIQKGELLENILSKLNIEDNKIILNNHTINFYGDINESRFS 1403
Db 1354 WNFVSYIVKSGKIEDNKPVLDPGFINNIPSTLKVSDNGFKGQFIS----IKNTPRAIN 1409
Qy 1404 LTFSELEDINIIEIDL-VSKSYKILLSGNCMKLIENSSD-IQQKIDHIGFNGEHQKYP 1461
Db 1410 LSFKINNI-VIVSYIHLNHSKNSITIISSDLNDIKNNFDNLDNINYLGLSISDNTIN 1468
Qy 1462 YSIDNETKNGFIDYSKKEGLFTAEFPNESIIRNYMPDSNNLPYSSKDLKDIRINK 1521
Db 1469 CIVRNDVYMEGKI-----FLNEK--KLVFIQNELELHLYDS-----VNK 1506
Qy 1522 GDVKLLIGN-----YFKDDMKVSLSPFTEDT-----NTIKLVGVLDENGVAQI 1565
Db 1507 -DSQYLNNPINNVVKYKDYIVGTEFLINSTENKYSIYIENNKIMLKLGLYL-ESSVFPT 1564
Qy 1566 LKFMNNAKSALNT-SNSLNFLESINIK---NIFYNNLDPN---IEFILDNTN-PIISGS 1616
Db 1565 IQDKIYSKEVNDYILSLIKKFFTVNIQCPMIVSGVDENNRYLEYMLSTNNKWIING- 1623
Qy 1617 NSIQGFELICDKKNQIQYFINKETSYTYLVGNRONLIVEPSYHLDGNSISSTVIN 1676
Db 1624 ---GYWE---ND-----FNNYKIVDF-----EKNVIVSGSNKLNSEGLDADT-ID 1662
Qy 1677 PSQKLYGIDRVNKKVIAPIAPLYT 1700
Db 1663 VLDKDLLENL--YIDSVIIIPKYYT 1684

RESULT 7

O68653

ID O68653 PRELIMINARY; PRT; 698 AA.

AC O68653;

DT 01-AUG-1998 (TREMELrel. 07, Created)

DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)

DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE Truncated toxin A.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 20309;
RX MEDLINE=99304497; PubMed=10376209;
RA Song K.P., Bai X.L., Chang S.Y.;
RT "Nucleotide and peptide sequences of the open reading frame encoding a
truncated toxin A gene of Clostridium difficile strain CCUG 20309.";
RL DNA Seq. 10:93-96(1999).
DR EMBL; AF053400; AAC08437.1; -- 021EB268A3BDEC5E CRC64;
SQ SEQUENCE 698 AA; 80682 MW; 021EB268A3BDEC5E CRC64;

Query Match 22.0%; Score 1911; DB 2; Length 698;

Best Local Similarity 52.9%; Pred. No. 7.6e-67;

Matches 371; Conservative 136; Mismatches 186; Indels 8; Gaps 5;

Qy 1 MVLVNAQLOKQWVVKERIQEDYVAIILNALSEYHNHNSSESVVEKYLKLDINNLTNKL 60
Db 1 MSLISKBELIKLAY-SIRPRENBYKTVLTNLDREYNKLTNNNENKYLQKKLNSIDVFM 59
Qy 61 NTYKSGRNKALKKFKKPYLTMVELEKNNSLTPVEKNLHFWIGQINDTAINYNOWKD 120
Db 60 NXYKNSRNRLSNLKKDLKEVILIKNSNTPVEKNLHFWIGVSDIVLEIKQWAD 119
Qy 121 VNSDYTVKVVDSNAFLINTLTKTIVESATNNTLESFRENLDPEFDYKFKRMEIY 180
Db 120 INAEYNKLMWYDSEAFVNTLTKAIVESSTTEALQLLEEBEQNPFQDNMKFKKRMFY 179
Qy 181 DKQHFIDYKSOIEENPEFIIDNIIKTYLSNEYSKDLKALNKYIEESLNKITANGNDI 240
Db 180 DRQRFNYYKSOINKPVTLTIDDIISYLVSEYNRDETLESYRTNSLRKINSNHGIDV 239
Qy 241 RNLEKFADEDLVRVLYNOELVERWNLAAASDLRLISMLKEDGGVYLDVILPGIOPDLFKS 300
Db 240 RANSLFTEQELLNINYNQELLNRGNLAAASDIVRLALKNFGGVYLDVDMFGIHSDLFKT 299
Qy 301 INKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNPMDLDEEVQVSFESALSCKSKSRIF 360
Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYNNTYSENFDFKLDQQLKONFKLIEESKSEKSF 359
Qy 361 LPDDIKVSPLEVKIAFANNNSVINQALISLKDSCSDVINQIKRKYKILNDNLPSINE 420
Db 360 SKJENLVNSDLEIKIAFALGVSINQALISKSGSYLTNLVIEQIKRKYQFLNQHLPATES 419
Qy 421 GTDFTNTWKIFSDKLASISNEDNMFMKIKITNLYKVGFPDVRSTINLSGPGVYTGAYOD 480
Db 420 DNNFTDTTKIFHDSLPNSATAENSFPLTKIAPYLOVGMPPEARSTISLSGFGASAYYD 479
Qy 481 LLMPKONSTNTHLEPELNRFPFKTKISQTEQIETSLWSFNQARAKSQFEYKKGYYE 540
Db 480 FINQENTIEKTLKASDLIEFKFPENNLSQLTEQEINSLSFSDQASAKYQFEKYKDYTG 539
Qy 541 CALGEDDNLDAQNTVLDKDY-VSKKILSS--MKTRKEYIHYIVOLQGDKISYEASCNL 597
Db 540 GSLSEDNEVDNPKNTALDKNYLNNKIPSNVVEAGSKNYVHYIQLQGDIDISYEATCNL 599
Qy 598 FSKDPYSILYQKNIEGSETAYVYVAD--AEIKEDKYRIPYOISNKRNIKLTPIGHGK 655
Db 600 FSKPKKSIITQRNW--NESAKSYFLSDGSGSIELELNKIRIPERKNEKVKVTFIGHGK 657
Qy 656 SEFNTDTFANLDVDSLSEIETILNLAKADISPKYIEINLL 696
Db 658 DEFNTSEFARLSVDLSLSNEISSFLDTIKLDISPKNVEVNL 698

RESULT 8

Q9EX08

ID Q9EX08 PRELIMINARY; PRT; 698 AA.


```

Db 541 LSENEVDNFKNVTDALDKNYLNNKIPNNVBERAGSKYVHYIIQGGDISYEATCNLFS 600
Qy 600 KDPYSSILYQKNIQEGSETAYYYVAD--AEIKEDIKYRIPYQISNKRNIKLTFIGHGKSE 657
Db 601 KNPKNIIQRMW--NESAKSHFLSDGSEILELNKYLIPERLKNKEKVKYTFIGHGKDE 658
Qy 658 FNTDTFANLVDVSLSEIETILNLAKAISPKEYIENLL 696
Db 659 FNTSEFARLSVDLSLSNEISSFLDTIKLIDISPKNVBEVNL 697

RESULT 10
Q9FCX5 PRELIMINARY; PRT; 553 AA.
AC Q9FCX5;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Toxin A (Fragment).
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C34;
RX MEDLINE=20402122; PubMed=10931294;
RA Braun V., Mehlig M., Moos M., Rupnik M., Kalt B., Mahony D.,
RA von Eichel-Streiber C.;
RT "A chimeric ribosome in Clostridium difficile combines features of
RT group I introns and insertion elements.";
RL Mol. Microbiol. 36:1447-1459(2000).
DR EMBL; AJ131844; CAC03681.1; --
FT NON_TER 553
SQ SEQUENCE 553 AA; 63913 MW; DB4551AGD8C3B25 CRC64;

Query Match 16.9%; Score 1469; DB 2; Length 553;
Best Local Similarity 51.7%; Pred. No. 9.1e-50;
Matches 287; Conservative 106; Mismatches 160; Indels 2; Gaps 2;

Qy 1 MNLVNAQLOKQVYKFIQDEYVAIINLALNEEYNNMSESSVVEKYKLKQINNTDNYL 60
Db 1 MSSIISKELIKLAY-SVRPRENEYKTLTNDLVEYNKLTNNENKYLQKLKLNESIDVFM 59
Qy 61 NTYKSGNKALKKEVILTWEVLELKNLSLTPVEKNLHFPIWIGQINDTAINYNQW 120
Db 60 NKYKNSRRNALSNLKQDILKEVILIKNSNTPVEKNLHFVWIGVEVSDIALEYIKOWAD 119
Qy 121 VNSDYTVFVDSNAFLINTLKKTIVESATNNTLESFRENLDPEFDYKFKRMEIY 180
Db 120 INAEYVNLWDSEAFVNTLKAIVESTTEALQLLEEIQNPQFDM-KFYKRMFIY 178
Qy 181 DKQHFIDYKQISENPEFIIDNTIKTYLNEYSKDLEALNKYIEESLNKITANNNDI 240
Db 179 DRQKRFINYKQINPKVPTIDDIKHLNVEYNRDETLLESYRTNSLRKINSNHGIDI 238
Qy 241 RNLEKFAEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPKS 300
Db 239 RANSLFTQELLNTYSQELNRGNLAAASDIVRLALKNFGGVYLDVDMPLPGIHSDLFKT 298
Qy 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEDEEVSFESALSKSKSEIF 360
Db 299 IPRSSIGLDRWEMIKLEAIMKYKYNINYSNFQDLQQLKONFKPLIIESKSEKIF 358
Qy 361 LPLDDIKVSPLEVKIAPANNVINOALISLKDSYCSDLVINOIKRNYKILANDNLPSINE 420
Db 359 SKLENLNSDLEIKIAFALGSVINQALISKQGSYVNLVIEQVKNRYQFLNQHLPATES 418
Qy 421 GTDFNTWKIPSDKLASISNEDNMFMFKITNYLVKGFAPDVRSTINISGPGVYTCAYQD 480
Db 419 DNNFTDTTKIFHSDLSFNATAENSFLTKIAPYLVQVGFMPPEARSTISLSGFCAYASYD 478

```

```

Qy 481 LLMFKDNSTNTHLLEPELRNFEPPKTKISQITSEITSLMSFNQARAKSFEYKKGYPF 540
Db 479 FINLQNTIETKLKASDLIEPKFPENNLSQLTQEINSLWSFDQASAKYQIERVYDVTG 538
Qy 541 GAIGEDDNLDPANQT 555
Db 539 GSPSGDNGVDNFNKT 553

RESULT 11
Q9ZGR4 PRELIMINARY; PRT; 3169 AA.
AC Q9ZGR4;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Putative cytotoxin.
GN I7095.
OS Escherichia coli O157:H7.
OC Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RX MEDLINE=983191744; PubMed=9722640;
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA Blattner F.R.;
RT "The complete DNA sequence and analysis of the large virulence plasmid
RT of Escherichia coli O157:H7.";
RL Nucleic Acids Res. 26:4196-4204(1998).
DR EMBL; AF074613; AAC70163.1; --
DR InterPro; IPR006473; YOP_T;
DR TIGRFAMs; TIGR01586; YopT_cys_prot; 1.
SQ SEQUENCE 3169 AA; 361971 MW; C2CDB48A72BDSFD1 CRC64;

Query Match 5.8%; Score 501.5; DB 2; Length 3169;
Best Local Similarity 20.1%; Pred. No. 3.8e-11;
Matches 421; Conservative 331; Mismatches 723; Indels 619; Gaps 104;

Qy 14 YVKFRI---QDEYVAIINAL--BEYNNMSESSVVEKYKLKIDNNLTDNYLTYKSGR 68
Db 160 YIKIRKTRGAEQDTTITQSLIINELLNGVDRTI-PPQKISELNDIHSYENMQIKNSR 218
Qy 69 N--KALKKFKYLTWEVLELKNLS-----LTPVEKNLHFPIWIGQ 106
Db 219 KGIBILVKQGSLLSLINDNKGKQLSDNASKIINLLGIYQSHKVDIEPPIHAWVAGA 278
Qy 107 INDTAINYINQWKNVSDYTVKPVYDSNAF-----LINTLKKT----- 144
Db 279 PPDNTFSYITAFNTYKDYTYLLWDIPNAGAAKPSGLKNIAINYAIRLRNTNPHLAE 338
Qy 145 -----IVESATNNTLE--SFRENLDPEPDY-----NKFYKRMETIYDQKHFI 187
Db 339 EMNEVLKIQNIQNETIEPKETRELRKLELNRYKSLTSETKEKNVFNLESWIGMHDNYF 398
Qy 188 DYKSKQISENPEFI-----IDNIK--TYLSNEYSKDLEALNKYIEESLNKITANN-- 238
Db 399 TYCISNGISNTDDISRLDFTLVNLKLSPEVQNDPKSTVEKNKRDIDLKNTISQKGDGF 458
Qy 239 ---DIRNLEKFAEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQP 295
Db 459 QLRDINTLESFKPKQDVFYQOQEMLLRNWYAAASQVRINILKEYGGIYTTDILPAYSD 518
Qy 296 DLFKINSKPSDITNTSWEMIKLEAIMKYK--EYIPG--YTSKNFDMLEDEEVSFESALS 351
Db 519 KVSQIINE-KSSDKRPFEDLKLRIISESILSLKGEKYSIKH-DGLDETTLNQLNNIL- 575
Qy 352 SKSKDKSEIFLPLDDIKVSPLEVKIAPANNVINOALISLKDSYCSDLVINOIKRNYK 395
Db 576 SEIEK---LTIDY-PKPVETKVVRTDFKIFKRYQKWNTWIRGNMNFMLTHGSKC 630

```


Db 856 LSTYNSTELNKLNFSSBKDDDELNVESKYKODENIFIEKNKIFDDIIDIKNYKNTAI 915
Qy 32 BEYHN-----NSESSEVVEKYKLKD--LNNL-----TDNLYNTYK-----KSGRNKAL 72
Db 916 KULNAININGSNNLSLSDVMKNGDINRLSQRSYLYQTDNFIDYIEKIFUKOWNLKGL 975
Qy 73 KKFKEYLTWEVL-----ELKQNSLTPVEKNL-----HFIW--IGQIND 109
Db 976 BEIENRLSNTYNNVELKIBAEQNEKYKXKENINTYDDTFLEKLIGDNYEWEVLKIELG 1035
Qy 110 TAINY-----INQWKDVSNDYTVKFVYDSNAFLNLTAKTIVESATNNTLESFR 158
Db 1036 LNVNTNLOANIDTTLIIRPYID-HIDHIIISLESKHNIEKNIKKIVIPNLERLKDFOIQTK 1094
Qy 159 ENLNDPEFDYNYFKRMEIIVDKQHFIDY-YKSOIRENPEFIDNIIKTLYLSNEYSKD 217
Db 1095 FNTNDIKLQHNLIIRI-----DNR-----DYNHMKLEEKEDLPKNI-----NDKKEE 1140
Qy 218 LEALNKYIEESLNK-----ITANNNGDI-----RNLEKFADE 249
Db 1141 TEKLKLEENNTKMENTNISVGIKKKNLIDYESMSSLKSIITDENLYKLQNV 1200
Qy 250 DLVR-LYNQELVERNWLAAASDILRISMLKEDGGVLOVDILP-----GIQDLFKSINKPD 305
Db 1201 DEFRLLYELILTEINEKIKQIKEASVERVEIELYKEKILSMNNSIKEDISK-----1254
Qy 306 SITNTSWEMIKLEATMKYKEYIPGYTKNPFMDLDEEVQSPESALSSKSKSEIFLPDLD 365
Db 1255 -LTPKYNEYDECITKENNIKELYT-XSSSLLE-----SCRDN-----MDI 1296
Qy 366 IKV--SPLE--VKIAPANNVINOALISLKDSCYSDLVINOIK-----NRYKILN 411
Db 1297 IKNKSVLDYELKTSIQONNDIKNSLASLKNYA---ILOSKLDVATKYILDNSYK--1351
Qy 412 DNLNPSINEGDTFNTMIFSKLASISNEDNMFMKIITNYLKVGPADPVRSTINLSGP 471
Db 1352 -----EDHARDLEMELESALSKN-----IK-----LKIEAAEYRNKV-----1386
Qy 472 GYVTGAYQDLMFKDNSTNIHLLEPELNFPEPKTKISQLTQEITSLWSP--NQARAKS 529
Db 1387 -----LGSEKHESIDY-----IRNIEKIQDISVIESDMTCIQNAYDNKTKML 1432
Qy 530 QPEYKKGY-FGALGEBDNLDPACTVLDKDYVSKILSSMKTKRNEYIHYIVLOQDK 588
Db 1433 HFQNVHRGDLILGNKN-----QGVIKPAESGNI-----SEY-----EQON 1472
Qy 589 IS-YEASCNLFK--DPYSSILYQKNEGSETAYYYYVADAEIKEDK-----633
Db 1473 INDYQOKCKTYSEASDNYEISYRKD-----SLLEFEKKITNILDVLI 1517
Qy 634 YRIPQISNKN-----IKLT-FIGHKSEFNTDTFANLVDVLSSEIETILNLA 684
Db 1518 FNMKTKLENKDSVNNILEGMKLTASIIINEKSHVLSSEKIRVURNQKNIKDBELLNNEKS 1577
Qy 685 DISPKVIEINLGCNMFYSIYAETYPCKLALKKD-RVSE-----LMPSI 730
Db 1578 KIAYQOFEL-----YMGKLYAIPDINRLHDKAERMFPSKAQFLMKPWL 1620
Qy 731 SODSITVSANOYEVINEBEGREILDHSGKWNKEESIIKOISKEYISFNPK-----N 785
Db 1621 DMSLILEKNLEELK-----NKEHDYMEI---EYIQDEBKHIKYELT 1660
Qy 786 KIIIVSKYLHELSTLQELIRNNANSSDIDLEKKVMLTECEINVASNIDROIVEGRIBAK 845
Db 1661 KLIEVIQFVDNIENTLEKRYKYEQGNLE-----NVYNNNSN--IKDRIEETK 1706
Qy 846 NLTDSINVKNEFKLIESISDLYDLKHQGLDSDHFSFISFEDISKTENGFRIRINKET 905
Db 1707 NLSVLVN-----IFSIKNSIYLKXHS-----VKMEDLNS-----1737
Qy 906 GNSIFETEKEIFSEYATHISKEISNIKOTIFPDNVNGKLKVKVNLDAAEHVNTLSAPFI 965

Db 1738 ----YIKOMNDIYDEF-----MESYNLLQKTIIESSN-----DDIEYEELKQVRNIEKFEI 1785
Qy 966 QSLTEYNTTKESNLNSLVAMKVQVYAOULFSGLNTITDASKVVELVSTALDETIDLLPTL 1025
Db 1786 QLV-----KAEDMKLYNSIKNDVYNKLVYIKWLELDNRCKSVDTILNEGDFPNKC 1841
Qy 1026 SEGPIIATIIDGV--SLGAAI-----KELSETNDPLLRQIEBAKIGAMVNIATASTAIVT 1080
Db 1842 AKESNVSGD-DDNVVYNELNKAINKYKEIHEKSNFVCKNEAESLFGIIVKSSNIIGMKIIT 1900
Qy 1081 SALGIASFSLIIVPLAGISAGISLVNNEL-----ILOQKATKYDVPKHI 1127
Db 1901 -GLG-----LELKEDVDLGTWLSLSSLHFHTASINKLYSTIESDVKVNDCLKS- 1950
Qy 1128 SLAETEGAFITLLDDKKIIMPQDDLVSEIDFNNSITLGCBIWRAEGSGHTLTDIDH 1187
Db 1951 SLDIVKYSFDEKKK--ROINSIMEEINNVHAHITI-KIE-----LANMINDS 1995
Qy 1188 FSSPSITVRKPLSIYDVLNIIKKEKIDFSKOLMVLPNAP-NRVFGYEMGW-----1236
Db 1996 RNKISVVLNKIYVSIKIKVVKEMTCDSSYHMIKMEGEYDKLKEYYKNYNEEKLATNE 2055
Qy 1237 --TGFPSRLNDGTLLDRIRDHYEGQFYWTFYAFIADALITKLKPRYEDTVNRINLQGN 1294
Db 2056 LNTNKLKENLSCKLLDELQNTMKTIEEB-----SPKKVIEDIKKSYDEINGRI---GN 2108
Qy 1295 TR--SPIVPVITTRQI-----RKNLSYSFVGS--1319
Db 2109 TEMDABAINVVEBELVRKQNRKASWYTSLLGKVNTKMSFKKLVLDRQKISQSFDSMK 2168
Qy 1320 --GGSYSLSLSPYNNIDNLVENDTWVIDVNVVKNITIESDEIQK-----ELIFENILS 1373
Db 2169 LNFNMINWINKINNHFKHQINN-----YSLGNVEKMMYINDSNEKTSVIELIDNLTK 2224
Qy 1374 KLN-----EDNKIILNNHTINPYGDNIS-----NRFLSLTFS 1407
Db 2225 KISINNSNNNNNNNNNNNNNDIINFINTKENLITNYENLGKCKXDLNIYLIKIFFS 2284
Qy 1408 ILEDINIIE-IDLVSKSYKILLSGNCMKLEN-----SSDLOQK-----I 1447
Db 2285 KLEEIEKSEKIFYSKAFNNVVKYQOKGLSDNINRMNINIKNIEEKETILNNINKNFTL 2344
Qy 1448 DHI-GFNGEHQKIFYYS---YIDNETKY-----NGFIDYKCKGLFTAESNESIIRNI 1497
Db 2345 DTIKEFNKAYDNILFHRDLVLEESKYDVGRKLNIIYIEY-----IYQIRNKSNNLINNF 2399
Qy 1498 Y-MPDSNNLFYSSKDLKD-----IRIIRNG-----DVKL-----LIGNYF---KDDMK 1537
Db 2400 YDIRNTGYIMIHETEKLNIDINEYMNKINKNINNTNKEFDITLRNKTNNYFVNNHBEIK 2459
Qy 1538 VLSFTIEDTWTIKLN-GVILDENGVAOILK-FMNNKASALNTSNLNFLESINIK-NI 1594
Db 2460 KVIDNFWNHLDTIKANFTFLOPEKVLVLIKSYLNDIK---NIMNELMR-KEQVDVSTOT 2515
Qy 1595 FYNLDPNIEPILDTNFIISGNSISIQFELICD-----KDKNIQYFYNFINKETS 1645
Db 2516 CTRNLIDIELEKI-----NKIKONNALTDDMVNLFSLKDKYKNEFLLTKDILKRV 2565
Qy 1646 YTLVGNRQNLIVBP-SYHLDDSGNISSTVINFSQKLYGIDRYVNVK 1692
Db 2566 QDKY--QEMNKIYRTLTVTNNDNEN-----KFALKYLSNVNAPFINEI 2605

RESULT 13
O82916
ID O82916 PRELIMINARY; PRT; 3169 AA.
AC O82916;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Toxin B.
GN TOXB.
OS Escherichia coli O157:H7.


```

QY 1436 LI--ENSSDIOQKIDHIGNGEHQKQYFYSYIDNETKNGFTIDYSK-----KEGLTAFBFS 1489
Db : : : : :
1802 LTFEGINSLSALEHMN-----IDAIMSVIGLVQYARMIKMKNNDNISAIIDHA 1848
QY 1490 NE-SIIRNIYMPDSNNLFYSSKDLKDIRINKGVK-LLIGNYFKDKMKVLSLTIEDT 1547
Db : : : : :
1849 GAVSDIKNIV-----DKFLGGILTLTNRRVYVPGVSGASLEGFTSSGLEVCAS----- 1897
QY 1548 NTKILNGV---YLDENGVAOILKFMNNAKSALNTSNLMNPLESINIKNIFYNLID---- 1600
Db : : : : :
1898 ---RMGCTAGRYL--SNVAKVIL-----PLLDI--GINWSLYDSSLNHAXA 1938
QY 1601 -PNIEFI---LPTNF-IISGNSISQOFE---LICDKDKNIQYFIFNKIKETSITLYV- 1650
Db : : : : :
1939 TTQIEVISTAIDVSPSSINTALSIGAIVPLAIA-----IVPITI-FSHEVKNAVYVN 1992
QY 1651 --GNFQNLIVPSYHLD-----SGNISSTVINFSQKLYG---IDRYVNVKVII 1694
Db : : : : :
1993 QINERHKLWLEAKYLDNGSAKVLINKATGIIDLSNNQVLNGLYIDMRENPPIL 2047

RESULT 14
Q8EMP8 PRELIMINARY; PRT; 3317 AA.
AC Q8EMP8;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Predicted cytoskeletal protein.
GN MYP1550.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN (1) SEQUENCE FROM N.A.
RP STRAIN=HP-2;
RC MEDLINE=22354719; PubMed=12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Tani T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL; AP004170; BAC43946.1; -.
KW Complete proteome.
SQ SEQUENCE 3317 AA; 385840 MW; 0641151BAPA992BB CRC64;

Query Match 5.7%; Score 494; DB 16; Length 3317;
Best Local Similarity 20.1%; Pred. No. 7.8e-11;
Matches 419; Conservative 355; Mismatches 701; Indels 610; Gaps 105;

QY 14 YVKFRIQ--EDYVAILNALEYHN-MSESSVVEKYL-----KLKD-----INNLTD 57
Db : : : : :
583 YINVRDLKLERQSKIANNMLEPIANKLSDGSVNEENFGQFEGRIKKIDDLISKIASERE 642
QY 58 NYLNTY-----KSGSRNKALK--KFKEY 78
Db : : : : :
643 NYENTYSSVILEVLANEAHKENVNLNLIKTEERKNLISBESKSAQNSLKLGELEL 702
QY 79 LTWE-----VLEKNSLTPEKVL-----HFWIGQNDTAINYNQWQD-- 120
Db : : : : :
703 IGLQNEEIESLDEKNQFISDYEQLLIEKREKLNDKIDKNSINDVLMNFNEKINDLE 762
QY 121 -----VNSDYTEVYDSNAFLINTLTKT-----IVESATNNTLESFRENLDPE 165
Db : : : : :
763 YSFDRLKNTINDSLDLATNENQKELFNKIEETLVNNVQIVKEEVKQDTQTILANNVELN 822
QY 166 FDNKPYRRMEIYDKQKHPIDYKYSQTEENPEFIIDNIKTYLSNEYSKDLALNKYI 225
Db : : : : :
823 EAYGSLNQCGW--TYDKRFNDID--SQPEQ-----FRYLLKQVTENL--INKNT 865
QY 226 E--ESLNKITANGNDINRLEKFADEDLVRLYNQBLVERWNLAASDILRISMLKEDGV 283
Db : : : : :
866 EYDEMNSIVNNISDNFKDLISGIEENVK-YSEDIK--NVAETNLNLRNKIGEFGEI 922

```

```

QY 284 YLDVILPGIOPDLFKSINKPDSITNTSWEMIKLEAIMKYKE---YIPGYTSKNPMDLDE 340
Db : : : : :
923 ---SEKISNLVADIENKKNKVAESQANTINYLETNLANQKELVGLIKNISSNDKFLDE 979
QY 341 EVQSFESALSKSDKSEIFLPLDDIKVSPLEVKIAPANNVINOALLISKDSYCSDLVI 400
Db : : : : :
980 VDKENHR-----LDNLKLNDELDELISLQNSEI--QALLIEKNDFIHDLEV 1022
QY 401 NQIKNRYKILNDNLNPSINEGTDFTNTMKIIFSDKLASIS-NEDNMMPMKI----- 450
Db : : : : :
1023 FLLEKKNFNNDENINELKQSAQEI---IELNNKIENTLEINPENRLEIDIHFSKSELG 1079
QY 451 -----TNY--LKVGFAPDVRSTIN-----LSG 470
Db : : : : :
1080 NQNDFFRNLEQOLTSNYEIIKNELNNDIVQVINNVNELSSVSVDRIENLSFDKFSKFD 1139
QY 471 PGVTVGAYQDILLMPKONSTNIHLLEPELURNEFPFKTKISQLT---EQBITSLWSNQARA 527
Db : : : : :
1140 IDIQFDSFRHLL--KDVTEEN-LIDKNSNYENINNVNNDLNTKFPESLITKLRSNNRTV 1195
QY 528 KSQPEEYKK-----GYFEGALGF--DDNLDFAQNTVLDKDYVSKKILS 568
Db : : : : :
1196 RLKNEIKKTTWRTNKKIDNSIIIEIGRFPANLIDIKQKNKF-ENSQLDVYV---SIE 1251
QY 569 SMKTRNKEYIHY-----IVQLQDKISYEASCNL-----FSKDPYSYIL 607
Db : : : : :
1252 ENKSHQREILAYIKNLENNNEFISQIEAQKFDNE--NLKLELEELISLQNEBIESIL 1308
QY 608 YOKNIEGSETAYVVYVADAETIKEDKYRIPQOISNRNIKL-TFIGHOKSEFN----- 659
Db : : : : :
1309 YEKN-----DFINDIQALLNKKK--KLDNENIIRLEDSINNAIGNFNPKISELE 1356
QY 660 ---TDTFANLDVDSLSEIETILNLAKADISPKYIEI-----NLGCMFSPSIVAEITY 711
Db : : : : :
1357 QNFNNKLANIDLSLSEED-----GKQVEFFDSQVNSLSHN---YELVKN----- 1400
QY 712 PGKLLIKIKRVSELMPSISQDSITV-----SANOVEVRINEGKREILD---HS 758
Db : : : : :
1401 -----PKDSLSLSILNLDVNTVENTNSKIGDISSEIRDISLNNQAKFDSITQDSD 1453
QY 759 GKWINKE--ESII-KDISKEYIS-----FNPKNKIIIVKSKYLHELSTLQIEIRNNANS 810
Db : : : : :
1454 FRHLLKEVTENLIDKNSNYENINNMDDINTRPESLITKL--NENRTVRKLNNEKCK 1511
QY 811 SDIDLEKKVMLTECEI-----NVASNDR---QIVEGRIEAKNLTSDSINYI 855
Db : : : : :
1512 TNLRTNKKLDNTISBIGDKFSDLIQEIKNQKNFESSQSDVIYSIIEENKSHQREILAYI 1571
QY 856 K-----NEF-----KLI-----ESISDSLYD-----LK 873
Db : : : : :
1572 KNLNEDNNNEFINQIEAQKNYNNENLKLKELEELISLQNEBIESLLEYKNDFINEVETPLN 1631
QY 874 HONGLDDSHFISFEDISKTENGFRIRFINKETGNSIFITETKEIPSEYATHI-----SKE 928
Db : : : : :
1632 KKNKLDNENINRLED---SVNNAIVFNKK-----ISALEENFNSRLANIDLSSLSEE 1681
QY 929 ISNIKDTITFDNV-----NGKLVKVNLDAAEV-NTLNSAFFIOSLLEYNTTKESLNL 981
Db : : : : :
1682 IDGQVEFFDSVQSLSYNYELVKNFEDKSLNLIENL-----EVTNVENVTNK 1732
QY 982 SVAMKVQV-----YAQL-----FSTGLNTITDASKVVELVSTALDETIOLLPTL 1025
Db : : : : :
1733 IADISTEIKDINLNNQSRFSDLSQFNFTLLKKEVTEG--PINKENSHYDELNLIISI 1790
QY 1026 SEGHP-IATIIDGVS-LGA-AIKELSETNPLRQIEIAKIGIMAVNLTAATAIVTSA 1082
Db : : : : :
1791 DESPKGLISSFKDSTNQIGENLVKDIKHSNKL---AIKMDSSIMEMN----- 1835
QY 1083 LGIASGFSILL-----VPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETGATFL 1138
Db : : : : :
1836 ---SKFDRLIEKMKVQTKDIENSVDTLN--MLIESKNKNQBEIFNTVSVLLE-KSTSSL 1888

```



```

QY 707 -----ABETPGKLLKIKDRVSELMPSISQDSITVANSQYVRINEGKR 752
Db 1303 SDIRKNSLKIIQDFSEESYINDIKLEKNVLE-----SQNNNT-DINQYLSKI--ENIY 1354
QY 753 BILDHSGWINKEESIIKDSSKEYISFNPKENKIIIVKSKYLHELSTLLOLRNNANSDD 812
Db 1355 NILK-----LNKIKKIIDKV--KEYTDEIRKNNK-----KINAEUS-----NSBKII 1394
QY 813 IDLEKKVMLTECEINVASNIDRQIVEGRIEAKNLTSDSIN-----YIK-----856
Db 1395 TOLKENSSLEKQOSKISTIDNNYVSECIKNITNLKTYIVNEKNWINTYFKNABEYNQNV 1454
QY 857 -----NEFKLIESISDLYDLKHQGLD--DSHFISFEDISKTENGFRIRFINKETG-NSI 909
Db 1455 SLNFNNIEMADTKSQYIILNKKNGTNTDYNIKELKEHKKKSNNYK-----DEAGKNTQ 1509
QY 910 FIETEKEIFSEYATHI-----SKEISNIKDTIFDNNVNGKLVKKNVLDAAHEVNT---958
Db 1510 EIKKNKELFEYEQEVTVLLNKYYAVELKNKFDKT-KNYSEQIIKEIK--DAHNTFTSQA 1566
QY 959 -----LNSAPFIQSLIEVNTTKESLSNLSVAMKVQVYAQLFSTGLNTITD-----1003
Db 1567 DKSEKKNMEIKNEQIRI EDEVAKNNKSNKAI-LDIQLSVPEPKIKFKIKDLRTKSDCL 1625
QY 1004 -----ASKVVELVSTALDETIDLTPLSEGLPIIATIIDGVSLGAAIKELSETND 1053
Db 1626 KETKDIEKISNLSIDTQETKLIENKNILNTLLEKLESKN--OKKNIEDQKKELDEVNS 1683
QY 1054 PLLROEIEAKIGIMAVNLTAATAIVTSALGIASGFSILLVPLAGISAGISLVNNELIL 1113
Db 1684 KI--KNIESNVNQHKNY-----EIGIVEKI-----NEIA- 1711
QY 1114 QDKATKVIDYFKHISLAETEGFTLLDDKIMPQDDVLSEIDFNNSITLGCBIWRAE 1173
Db 1712 --KANK-----DOIESQKLIITIKNLISPPKANDLEGIDTWN---LCK---YNTY 1756
QY 1174 GSGH-----TLDDIDHFFSPS---ITVRKPWLSIYDVNLNKKEDIFSCKDLMLVLPNAP 1226
Db 1757 MNVIEEFIKSYDLITHVLETVSKEPIVTEQ-----IKNKRIITAQNELLTNKNV 1806
QY 1227 NRVOGYEKGWTPGFRSLONGDKLLDRIDRHYEG-----QFYWRY-----PAFIADA 1273
Db 1807 NKAKSY-----LDDIEANEFDRIVTHFKNKLNDVNDKFTNEYSKVKNKGFNDISNS 1856
QY 1274 LITKLPRVEDTNVRINDGNTRSPVIVITTE-----OIRKNLSYSFY 1317
Db 1857 -INNVK-KSTDENLNLINLQTKENYANIVSKYYSKYAEARNIFINIPKANSNIQIK 1914
QY 1318 GSGG-----SYSLSLSPY---NMNIDLNLV-----ENDTWVIDVNVKNITTESDEI 1362
Db 1915 SSSGIDLFPKNINIALPYLDSQKDTLTFIPSPKTSYTYTKISDSYNTLLDLKRSQEL 1974
QY 1363 QKGE--LIENILSKLIENKIIILNHTINFYGDI--NESNRFISLTFISLEIDNIIIEIDL 1420
Db 1975 QKKEQALNLIPENRLLHDVKQATNELKDTLSDLNKKKEQILNKKLLHKSNEMLNKLSC 2034
QY 1421 VSKSY-KILLSNCMKLIENSSDIOQKIDHIGFNGEHQYIFYSVIDNETKNGFI-DYS 1478
Db 2035 NSQYDTLLESKYDKIKEKSNYKEKENLGIN-----FDVKAEEQFNNDIKOIE 2086
QY 1479 KKEGLFTAEFNESIIRNIYMPDNNFLIYSSKOLKDIRIINKGDVKKLLIGNIFYKDDMKV 1538
Db 2087 KLENNY-----KHSEKDNVNFSEENNILQSKKKLKELT-----NAFNAEIK- 2128
QY 1539 SLSTFIETNTIKNGVLDENGVAQILKFPNNAKSA-LNTSNSLMNFLESINIKNIFYN 1597
Db 2129 -----KIBDKIIIEKNGL--INKLIETKDXCMLFTYKTL---VETLKIKTIDY- 2170
QY 1598 NLDPNIEFILDFTNFIISGNSISQFELICDKDKNIQPYFINKETSYTLVVGNRQ--N 1655
Db 2171 -----TKFTITSATKFSKFKLYDATSNLNDIN--TLQTKYDLNQLNKHVAS 2217
QY 1656 LIVEPSYHLDOSGNI-----SSTVINFSQKLYGIDRYNVNKV 1692

```

```

Db 2218 MVADAT---NDNNNLIEKEKEATKTINNLTJ--LFTIDS--NKI 2254

Search completed: November 5, 2003, 19:31:48
Job time : 59 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:25:47 ; Search time 15.625 Seconds
(without alignments)
5116.506 Million cell updates/sec

Title: US-09-126-816B-6_COPY_1_1700

Perfect score: 8677

Sequence: 1 MNLVNAQLQRMVYKFRQ.....YLYGIDRVYKVIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result; being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6733.5	77.6	2366	1 TOXB_CLODI	P18177 clostridium
2	4458	51.4	2710	1 TOXA_CLODI	P16154 clostridium
3	351	4.0	2869	1 RBP1_PLAVB	Q00798 plasmodium
4	317.5	3.7	1726	1 MSP1_PLAPC	P04934 plasmodium
5	316.5	3.6	1701	1 MSP1_PLAPF	P13819 plasmodium
6	315.5	3.6	1682	1 MSP1_PLAP3	P13998 plasmodium
7	313.5	3.6	1701	1 MSP1_PLAPF	P08569 plasmodium
8	312.5	3.6	1726	1 MSP1_PLAPF	P50495 plasmodium
9	309	3.6	1639	1 MSP1_PLAPF	P04933 plasmodium
10	306.5	3.5	2136	1 YCP2_MARPO	P09975 marchantia
11	306	3.5	1630	1 MSP1_PLAPK	P04932 plasmodium
12	305.5	3.5	1957	1 SPOF_SCHPO	Q10411 schizosacch
13	301.5	3.5	1251	1 RBP2_PLAVB	Q00799 plasmodium
14	292	3.4	2748	1 NUM1_YEAST	Q00402 saccharomyc
15	290	3.3	1956	1 MLP1_YEAST	Q02455 saccharomyc
16	289	3.3	1956	1 ATX1_PLAPF	Q04956 plasmodium
17	287	3.3	3433	1 UTRO_HUMAN	P46939 homo sapien
18	280.5	3.2	2059	1 TEGU_HSV7J	P52362 human sapie
19	278.5	3.2	6885	1 SNE2_HUMAN	Q8wxh0 homo sapie
20	274	3.2	2339	1 RCP1_PLAPF	P27625 plasmodium
21	273.5	3.2	1928	1 MY51_YEAST	P08964 saccharomyc
22	272	3.1	5171	1 BPEA_HUMAN	O94833 homo sapie
23	269.5	3.1	1892	1 R835_RICCN	Q92hd6 rickettsia
24	269	3.1	1803	1 YJL3_YEAST	P47024 saccharomyc
25	268.5	3.1	2167	1 BEN2_YEAST	P39960 saccharomyc
26	268	3.1	2663	1 CENE_HUMAN	Q02224 homo sapie
27	267	3.1	1679	1 Y109_YEAST	P40457 saccharomyc
28	265.5	3.1	1727	1 ALM1_SCHPO	Q9utk5 schizosacch
29	262.5	3.0	4092	1 DYHC_YEAST	P36022 saccharomyc
30	261.5	3.0	1628	1 NAGH_CLOPE	P28631 clostridium
31	261.5	3.0	3210	1 CENF_HUMAN	P49454 homo sapie
32	260	3.0	2014	1 YJU7_YEAST	P39526 saccharomyc
33	259	3.0	4196	1 DYHC_SCHPO	O13290 schizosacch

RESULT 1

ID	TOXB_CLODI	STANDARD;	PRT;	2366 AA.
AC	P18177;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Toxin B.			
GN	TOXB OR TCDB.			
OS	Clostridium difficile.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1496;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VPI 10463;			
RC	MEDLINE=90325540; PubMed=2374729;			
RA	Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;			
RT	"Nucleotide sequence of Clostridium difficile toxin B gene."			
RL	Nucleic Acids Res. 18:4004-4004(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VPI 10463;			
RA	von Eichel-Streiber C.;			
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1271-2366 FROM N.A.			
RC	STRAIN=VPI 10463;			
RC	MEDLINE=92293124; PubMed=1603068;			
RA	Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S., Schulze J.,			
RA	Sauerborn M.;			
RT	"Comparative sequence analysis of the Clostridium difficile toxins A			
RT	and B.";			
RL	Mol. Gen. Genet. 233:260-268(1992).			
CC	-1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN ENTEROTOXIN			
CC	CAUSED A AND CYTOTOXIN B.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X53138; CAA37298.1; -			
DR	EMBL; X92982; CAA63562.1; -			
DR	EMBL; X60984; CAA43299.1; -			
DR	PIR; A27636; A27636.			
DR	PIR; S10317; S10317.			
DR	InterPro; IPR002479; CW binding.			
DR	Pfam; PF01473; CW binding_1; 17.			
DR	Pfam; PF04488; Gly_transf_sug; 1.			
KW	Toxin.			
SQ	SEQUENCE 2366 AA; 269709 MW; E1024BD8B8A56ADF CRC64;			

34 257.5 3.0 1162 1 BXEN_CLOBO P46082 clostridium
35 257.5 3.0 3135 1 S230_PLAFO Q08372 plasmodium
36 255.5 2.9 2875 1 RRPL_TSWV1 P28976 tomato spot
37 255 2.9 1790 1 US01_YEAST P25386 saccharomyc
38 254 2.9 3144 1 VF13_YEAST Q07878 saccharomyc
39 253 2.9 1295 1 BXAL_CLOBO P10845 clostridium
40 251.5 2.9 8545 1 ANCI_CAEEL Q9n4m4 caenorhabdi
41 250.5 2.9 3305 1 APLP_MANSE Q25490 manduca sex
42 250 2.9 4563 1 APB_HUMAN Q04114 homo sapien
43 249 2.9 1274 1 BXF_CLOBO P30996 clostridium
44 248 2.9 1276 1 BXD_CLOBO P19321 clostridium
45 248 2.9 3259 1 GIAN_HUMAN Q14789 homo sapien

ALIGNMENTS

RA Wilkins T.W., Johnson J.L.;
 RT "Molecular characterization of the Clostridium difficile toxin A
 RL gene.";
 RL Infect. Immun. 58:480-488(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VPI 10463;
 RA von Eichel-Streiber C.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC 1- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
 CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
 CC DIFFERENT OLGOPETIDES.
 CC DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
 CC ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL
 CC DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE
 CC CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X51797; CAA36094.1; -;
 DR EMBL; M30307; AAA23283.1; -;
 DR EMBL; X92982; CAA63564.1; -;
 DR InterPro; IPR002479; CW_binding.
 DR Pfam; PF014703; CW_binding_1; 28.
 DR Pfam; PF04488; Gly_transf_aug; 1.
 KW Toxin; Enterotoxin.
 SQ SEQUENCE 2710 AA; 308052 MW; 0A6B52CE84C14421 CRC64;

Query Match 51.4%; Score 4458; DB 1; Length 2710;
 Best Local Similarity 50.7%; Pred. No. 3.7e-162;
 Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;

QY 1 MNLVNAQOLQKVVYKFRIOEDYVAILNALREYHNMSSVVEKYKLDKLNLTNDYL 60
 DB 1 MSLISKEELIKUAY-SIRPRENEYKITLNDYKLNLTNNENKYLQKLNESIDVPM 59
 QY 61 NTYKSGRNKALKKPKFVLTMEVLKNNSLTPVEKKNLHFVIGQINDTAINYNQWD 120
 DB 60 NKYKTSRRNALSNLKCOLKEVILKNSNTSPVEKNLHFVIGEVSDIALEYIKQWAD 119
 QY 121 VNSDYTVKPYVDSNAFLINTLKKTVESATNTNLTSEFRENLDNPDYNNKFRKRMETIY 180
 DB 120 INAEYNIKLYWDEAPLVNLTAKAIVESSTTEALQLEBEIQNPQFDNNKFKYKRMETIY 179
 QY 181 DKQKHFIDYKSOIBENPFIIDNLIKTYLSNEYSKOLEALNKYIEESLNKITANNNDI 240
 DB 180 DRQKRPINYYKSOINKPVTPTIDDIKSHLVSEYNRDETVELSSYRTNSLRKINSNHGIDI 239
 QY 241 RNLEKFAEDLRLNQLVERWNLAASDIIRISMLKEDGGVYLDVILPGIQDLPFKS 300
 DB 240 RANSLFTEGELLNYSQELLNAGNLAAASDIYVLLALKNFGVYLDVDMPLGHSDFKT 299
 QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVORFESALSSEKSKSEIF 360
 DB 300 ISRPSIGLDRWEMIKLEAIMKYKYYNNYTSNFENFDKLDQQLKDNFKLIESKSEIF 359
 QY 361 LPLDDIKVSPKVKAFAPANNVINQALISLKDSCYSDLVINOIKRYKILNDNLNPSINE 420
 DB 360 SKLENLNVSDLEIKAFALGSVINQALISKQSYLNLNVLIEQKRYQLNQLHAPAIAS 419
 QY 421 GTDFNTNTMKIFSDKLASINEDNMFMKITYNLKVGFAPDVRSRTINSLGPGVYTGAYD 480
 DB 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQGFMPPEARSTISLGPAYASAYD 479
 QY 481 LLMFKDNSTNIHLLBEPANFPFKTKISQLTEQEIITSLWSFNQARAKSQFEYKKGYPE 540
 DB 480 FINLQENTIEKTKASDLTEFKFPENNLSQLTEQEIINSLWSFDQASAKYQFEKYKRYDYG 539

QY 541 GALGEDDNLDAQNTVLDKDY-VSKKILSS--MKTRNKEYIHYIVLQGDKISYEASCNL 597
 DB 540 GSLSEDNGVDNKTALDKNYLLNNKIPSNVVEEAGSKYVHYIIQLQGDISIYEATCNL 599
 QY 598 FSKOPYSILYQKNI EGSETAYYYVAD--AEIKEDIKRYPYQISNKNKILTFIGHGK 655
 DB 600 FSKNPKNSIIIQRM--NESAKSYFLSDGSGSILELNKRYIPERLKNKEKVKVTFIGHGK 657
 QY 656 SEENTDTFANLDVDSLSEIETILNLAADISPKYIEINLLGCNMFYSIYAEETYPGL 715
 DB 658 DEFNTSEFARLUSDLSNEISFLDTIKLIDISPKNVEVNLGCNMFYSYDFNVEETYPGL 717
 QY 716 LLKIKDRVSELMPSISQDSITVSNQYEVRIINEGKREILDSHGKWKINKERSIIDKISS 775
 DB 718 LLSINDKITSITLPDVKNKSIITIGANQYEVRIINSEGRKELLASHGKWKINKERAIMSDLSK 777
 QY 776 EYISNPKNKIIVKSKYLHBLSTLQIBIRNANSSDIIDLEKVMMLTCEINVANIDRQ 835
 DB 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLALLDASVSDTKFTLNNLKLNISSIGDY 837
 QY 836 IVEGRIEAKNLTSDSINYKNKFKLIESISDSLYDLAKHQCLDDSHFISPEDISKTEG 895
 DB 838 IYEEKLEPVKNI IHNISDIDLIDEFNLLNENSDLEYELKLANLDEKYLISFEDISKNNST 897
 QY 896 FRIRINKETGNSIFETEKEIFSEYATHISKEISNIKDTIPDNVNGKLVKKNVLDAAHE 955
 DB 898 YSVRFINKNGESVYVEKEIFSKYSEHITKEITKNSIITDVNGNLLDNQLDHTSQ 957
 QY 956 VNTLSAFFIOSLIEYNTTKSSLSNLSVAMQVYQALFSTGLNTITDASKVVELVSTAL 1015
 DB 958 VNTLNAAFFIOSLIDYSSNKDVLNLSVSVKQVYQALFSTGLNTIYDSIQVLNLSNAV 1017
 QY 1016 DETDLLPTELSGLPIIATIIDGVSLGAAIKELSETNDPLLRQETEAIGTMAVNLTAAS 1075
 DB 1018 NDTINVLTPIITTEGPIVSTILDGILNGAAIKELDEHDPLEKLEAKGVGLVAINMSUSI 1077
 QY 1076 TAITVSALGIASGFSILLVPLAGISAGIPSLVNNELIQLDKATKVIDYFKHISLAETGA 1135
 DB 1078 AATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELIHLDRATSVVNFHLSSEKGYG 1137
 QY 1136 FTLLDDKIIMQDDLVLSEIDFNNNNSITLKGCEIWRAGGGSHLTDDIDHFFSPSITY 1195
 DB 1138 LKTEDDKILVPIDLVISEIDFNNNNSIKLGTNCLAMEGGSGHTVTGNIDHFFSPSIS 1197
 QY 1196 RKPWLSIYDVNLNIKEKIDFSLKOLMVLNPNRVPFGYEMGWTGPGERSLDNDCGTLDR 1255
 DB 1198 HIPSLISYSAIGIETENLDFSKKNMMLPNAPSRVFWETGAVPGLSLENDGTRLLDSIR 1257
 QY 1256 DHYEQFYRYFAFTADALITKLPYEDTNNRINLDGNTRSFIVPVITTEQIRKNLSYS 1315
 DB 1258 DLYPGKFWRYFAFF-DYAITTLKVEYEDTWIKIKLDKDRNFIMPTITTNIRKNLSYS 1316
 QY 1316 FYGGGYSLSLSPYNNMIDNLNVENDTWIDVNVNKNITIESDEIOKGLIENILSKL 1375
 DB 1317 FDGAGGTYSLLSYSPITNINLSKDLWIFNIDNEVBISIENTGTIKKGLIKDLVLSKI 1376
 QY 1376 NIEDNKIILNHNTHNFYDINESNRPISLTFSILEDINIIEIDLVSYSKILLSGCNCK 1435
 DB 1377 DINKNKLIGQITDFSGDIDNKKORYIFLTCELDDKISLIIEINLVAKSYSLLSGDKNY 1436
 QY 1436 LIENSSDIQQKIDHIFNGEHQKIFYFYVID-NETKYNGFDYDKKEGLFTAEFNSNII 1494
 DB 1437 LISNLNTEIKINTLGLD---SKNAYNYTDESNNKYFGAI-----SKTSQKSI 1483
 QY 1495 RNINMPDSNNL-----FIYSSKDL--KOIRINKGDKVLLIGNYFKD---DMKVSLS 1541
 DB 1484 H--YKDKSKNILEFYNDSTLEFNSKDFIAEDINVFEMKODINTITGKYVYDNNNTKSIDPS 1541
 QY 1542 FTIETNTIKLVGVLDENGVAQIILKPMNNAKSALNTSNLMNPLESINKIIFYNNLDP 1601
 DB 1542 ISLVSKNQVKNGLYNESVSYSDYDFVKNSDGHENTSNMNLFLDNISFWKLPGFE--- 1598

QY 1602 NIEFILDITNFIIGSGNSIGOFELICDOKNIOPIYFNFKIKETSVTLVYGNRONLIVEPS 1661
 Db 1599 NINVIDKIFTVGTNLTGVFEFIDONNKNIDIIYGEWKTSSKSTIFSGNGRNVVVEPI 1658
 QY 1662 YHLDSDSGNISVINFSQKYLGYDIDRYVKNVIAIPNLTY 1700
 Db 1659 YN-PDTGEDISTLSDFSYEPLAGIDRYINKVLIADPLTY 1696

RESULT 3
 RBPI PLAVB
 ID RBPI PLAVB STANDARD; PRT; 2869 AA.
 AC Q00798;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 GN Reticulocyte binding protein 1 precursor.
 DE RBPI.
 OS Plasmodium vivax (strain Belem).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 merozoites";
 RL Cell 69:1213-1226(1992).
 CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 CC HUMAN RETICULOCYTE CELLS.
 CC -!- SUBUNIT: Homodimer (Potential).
 CC -!- SUBCELLULAR LOCATION: Membrane-bound.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M88097; AAA29743.1; -
 KW Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
 FT DOMAIN 18 2807 EXTRACELLULAR.
 FT TRANSMEM 2808 2826 POTENTIAL.
 FT DOMAIN 2827 2869 POTENTIAL.
 FT SITE 1030. 1032 CYTOPLASMIC.
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
 SQ SEQUENCE 2869 AA; 330213 MW; 89DBE442205BECFF CRC64;

Query Match
 Best Local Similarity 19.7%; Score 351; DB 1; Length 2869;
 Matches 399; Conservative 323; Mismatches 633; Indels 626; Gaps 105;
 QY 6 KQALQKQVYVVKFRIQDEY-----VAILNALEBYHNMSSSVVEKYLK-LK-DINNL- 55
 Db 946 KALKEIVDSLRDKIDQYETEFKEKTSVAVENTVSTIQSLSKA--IDSLKRLNGSINCK 1003
 QY 56 ---TD-NYLTNVKSGRNKALK-----KFYEYLTMEVLELKNLSLTPVEKNLFIWIG 104
 Db 1004 KYNTDILRLSKIKTLUREEVQKMPKRGDKCGENTALLKLSLRDMKMGKINEKLN-----D 1059
 QY 105 GOIN--DT-----AINYINQWKDY-----NSDYTV- 127
 Db 1060 GRNLSLDTKKEDLLKFSYSESKSIHLSKDKGQDPQPLNRIDSEWDIKRVDLSLVNYYQVI 1119
 QY 128 -----KFVYDSNAFL-----INTLKTIVISATNTLESFRE-----NLNDPEFDYNK 170
 Db 1120 SENKVTLPFNKSVTYTEAMHSHINTVAHGIT-SNKNEILKSVEKVEDKLNLVQEONEDYKK 1178
 QY 171 F----YRKMEII---YDQKHFIDYKQIEENPEFIIDNIKIYLSN-----EYSKDL 219

Db 1179 VKNPENKQLEAIRGSMKSLKEVINKHVSEMTQ-----LESTANTLKSNAKGNEHDL 1233
 QY 220 ALNKY-----IESLANKI-----TANNNDIRNLEKFADEDLVLYNQELVERWNLA 266
 Db 1234 ELNKTGQMRDIYEKLAKIABELKEGTWELKDANERKANKVEPEPERNIIGHVLE----- 1288
 QY 267 AASDILRISMLKEGGVYLDVILPGIQPOLFKSINKPDSITNTSWEMIKLEAMKYKEY 326
 Db 1289 -----RITVEKQAGV--VEEMNSLKTKEIQE-----TSDDSQNELVTTSTIKLEN 1337
 QY 327 IPGYTSKNFMDLDEVORSFESALSSKSDKSEIFLPLDDIKVSPLEVKI-----AFANNSV 382
 Db 1338 AKGY-----EDVIKNEEDSIQLR-EKAKSLETIDEMKLVQVNNMIQSIQGNAG 1388
 QY 393 INQALISLSDSYCDLVINQIKRYKILNDNLNPSINEGTDFNTMTKIFSKLA----- 436
 Db 1389 ISKEL-----NELKGVIELLISYSSILLEYVKKNSSSVRFPSQLANGEFTK 1435
 QY 437 SISEDNMFMFKITNYLVKVGFAFDVRSTINLSGPGVYTGAYQDLLMFKDNSTNIHLEP 496
 Db 1436 AEGEKNASARLAEAEKLEQIVKDL-----DYSD-----IDD 1469
 QY 497 ELRNFEPKTKISQLEITSLMSFNQARAKSOFEEYKKGYPFEGALGEDDNLDAQN-- 554
 Db 1469 KVKIEGKREILKMKESALT-FWESE-----KFKQMCSSHMEKAKGKKKIEYKNG 1522
 QY 555 -----TVLDK-----DYVSKILSMKTRNKRYHYIYQIQGDKISYEASC----- 595
 Db 1523 DGKANITDSQMEEVGNVSK-----AEHAFHTVEAQVDKT--KAFCEISIVAVYT 1570
 QY 596 ---NLFSKDPYSSILYQKNIEGSETAYYYYVADAEIKEDIKYRIPYQIS-NKRN- 646
 Db 1571 KQDNLFNESLMKVKVCKEKNDEAEKY-----SAKLKPYDG-RIKARVSENERKISELKE 1625
 QY 647 KLTFIGHGKSFN-TDFANLVDVSLSEIETILNLAADISPKYIEINLGCNMFYSI 705
 Db 1626 KAKYEKESOLDVSTYKSLQIDNCQQLDSVL-----NIGRVKQNALQYFD 1674
 QY 706 YAEPTYPKLLKTKDRVSELMPSISQDSITVSANQYEVRIENERGKREILDHSGKWINK 765
 Db 1675 SADKSMKSVL-----PISLGAESLKDVKAAKESYEKNL-----ETVQNMESRINVE 1722
 QY 766 ESIKIDISSKEYISFNPKENKIIVKSKYLHELSTLQEIERNNANSDDIDLEKKVMTCE 825
 Db 1723 EGSUTDIDKK-----ITDIENDLLKMKQYEE--GLLQKIKENADKRSNFE-----LVGSE 1772
 QY 826 INVASNTDROI-VEGRIEEAKNLTSDSINY-----IKNEPKLIESDSLYDLKHQNGL 878
 Db 1773 INALLDPSTSIKLUKE-YDMTGDLKQYGVKNEIHGEF-----TKSYNLIETHLSNAT 1827
 QY 879 DSHFISPE-----DISKTENG-----FRIRPINKETGNSIFITEKEIFSYA 922
 Db 1828 DYS--VTPEKAQSLRELAEEHEHLRRREBAIFLLANDIKKVESLKLKEMMKVKSABYE 1885
 QY 923 -----THISKEISNIKDTIEDNVNG-----KLVKCV-----NLDAHEVN 957
 Db 1886 GMRKDRHVSQVQDMK-TIVDELKTLNDISECSSLVNNVSVIYKVKVSKHADYRRDAN 1944
 QY 958 -----TLNSAFF-----IOSLIEYNTTKESLNSLVAMKVQVYAOLFSTGLNTITDAS 1005
 Db 1945 SMYSEWVTLANYFLSDEAKISSGMFNA--EMKSNFKTDLEIF-----SVISNSN 1994
 QY 1006 KVELVSTALDETIDLPITLSEGLPIIATIDGVSLSGAAIKELSETNDPLRQETEAIG 1065
 Db 1995 ELKKIEQDSNDVIQ-KERESEQLAKADTIYV-----IKLKNFEKEL--EEAKNKE 2046
 QY 1066 IMAVNLTRAATVITSALGIASGILLSVPLAGISAGIPSLVNNELIILQDKATKVIDYFK 1125
 Db 2047 VVSEKVRALKRL-SQVEGIRCHF-----NPHRLDN-----TELENLK 2086
 QY 1126 HISLAETEGATLLDDKIIIMPQDOLVLEIDFN--NNSITL-----GKCEIWRAE 1173

Db 2087 KM-----VTIYDKKSERESGLQEMENNTYSNSITQLEGIVSAGESKEDIKLE 2138
 QY 1174 GSGH-----TLTDDIDHFFSSPSITYR-----KPMLSIYD-VLNKKEKIDFSK 1217
 Db 2139 RSNEWRNISEKISIDSQVEMWSTIDELYGKQKQAHWLSLSYANMKTSK---K 2194
 QY 1218 DLWVLPNAPNRVFGYEMGWTPGFRSLNDGTLKLRDRDHVBCQFYWRYPAFIADALITK 1277
 Db 2195 LIMINKENT-----EKVDYIKDNSSS-----TDGYVET 2225
 QY 1278 LKPRVETNVRNLGNTRSFVPVITTEQIRKNSLSYSPYGGSGSYLSLSPYNNIDLN 1337
 Db 2226 LK-----GPIGSKLTFSSA-----SE 2241
 QY 1338 LVEN-DTWIDVDNVNKNITTESDEIOXGELIENILSKLNIEDNKILNHNHTINFGDIN 1396
 Db 2242 IVQADTVSNVPAKESKSLNAIRDIKELYLFHONSISIVEGGV---QNMALYDKLN 2298
 QY 1397 ESNRFISLTSILEDINIIIEIDLVSYSKILLSGCMKLIENSDDI-QOKID-HIGFNG 1454
 Db 2299 BEKR-----ENDELVRN---ISETKLKQMEHSTDVFKPMIELHKGWNE 2338
 QY 1455 EHQQYIFYSIDNETKNGFDISK-----KEGLFTAEPNESIIRNIYMPDNNULFI 1507
 Db 2339 TNKK---SLEKEKKLSVNDHMSMEAEIMKNGL---KYTPES-VQNI-----NNIYS 2385
 QY 1508 YSSDKLDKIRIINK---GVKLLIGNY-FKDDMKVSLSTIEDTNTIKLNGVYLDENGVAQ 1564
 Db 2386 VTEAEVKLEEDRDYGD-----NYQIVBEHKQFSLIDRTNAL-----MDD---IE 2430
 QY 1565 ILKFMNNAKSALNTSNLMNPFLESINIKNIFFYNNLDPNIEPTLDTN-FIISGNSIGO-- 1621
 Db 2431 IFKKN-----YNLMEVATETIHRVNDYIEKTNKLVQAK 2466
 QY 1622 --FELICD-----KDKNIQYPIFNFKIKETSY--TLVGNRQNLIVPEPSYH 1663
 Db 2467 TEYEQILENIKQNDMLQNIPLKVSITIEYFENVKKKESILNDLY---EQERLLKIGB 2523
 QY 1664 LDD-SGNISSTVINP--SQKY-----LYGIDRYNKKV 1692
 Db 2524 LDEIKRNVETLSSYEIDQKMEMMSKNLEKSKRMVNTSYIELBREANEI 2574
 RESULT 4
 MSPL_PLAFC
 ID MSPL_PLAFC STANDARD; PRT; 1726 AA.
 AC P04934;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMNSA) (P195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Camp / Malaysia).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=5835;
 RN [1]
 RP SEQUENCE OF 1-1103 FROM N.A.
 RX MEDLINE=86205236; PubMed=3517809;
 RA Weber J.L., Leininger W.M., Lyon J.A.;
 RT "Variation in the gene encoding a major merozoite surface antigen of
 RT the human malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 14:3311-3323(1986).
 RN [2]
 RP SEQUENCE OF 1104-1726 FROM N.A.
 RX MEDLINE=88143999; PubMed=3278296;
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
 RT "Merozoite surface protein sequence from the Camp strain of the human
 RT malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 16:1206-1206(1988).
 CC -1- SURCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (potential).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42

kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X03831; CAA27446.1; -.
 DR PIR; A23386; SAZQGM.
 DR InterPro; IPR006209; EGF_like.
 DR Pfam; PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 FT Transmembrane; GPI-anchor.
 FT SIGNAL 1 19
 FT CHAIN 20 1726
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT SEQUENCE 1726 AA; 196197 MW; D88AD45FA3528CF3 CRC64;
 SQ
 Query Match 3.7%; Score 317.5; DB 1; Length 1726;
 Best Local Similarity 19.9%; Pred. No. 6.2e-05;
 Matches 350; Conservative 284; Mismatches 602; Indels 523; Gaps 89;
 QY 51 DNNLTNYLNTYKSGRNKALKKFKVLTWEVLKNSLTPVEKNLHFIWIGQINDT 110
 Db 154 DLKRVNRYLT-----IKELKYPELFDLTNMLTLCN-NTH----- 189
 QY 111 AINY-INQKDVNSDYTVKPVYDSNAFLINTKKTIVESATNTLE-SFRENLDPEPDY 168
 Db 190 GPKVLIDGYEIN-----ELLYKLN-FYFDLLRAKLVNDVCANDYCOIFNLKIRANELDV 243
 QY 169 NK----FYRKMEIYDKQHFIDYKSOIEENPEFIIDNTIKTYLSNEYSKOLBALNKY 224
 Db 244 LKLVFGYRKPLDNIKNVGMEDYIKKN-----KTTIAN-----INEL 282
 QY 225 IEESLNTKTNNGNDIRNLEK---FADEDIYRLYNQELVERVNLAAASDILRISMLKEDG 281
 Db 283 IEGSKTTIDQNKADNEEGKKLYQAQYDL-SIYNKQLEEAHNLISVLE-KRIDTLKKN 340
 QY 282 GVLVDVILPGIO-PDLFKSINKPDSINTSWEMIKLEIMKYKEYIPGYTSKNFMDLDE 340
 Db 341 NIKELLKINIKVPPPPANSNGTPTNLDKKNKKEEHEE--KIKE-IAKTIKFNIDSFT 397
 QY 341 EVQSFESALSSKDSKSEIFLPDDIKVSPLEVKIAFANNSVINQALISLKSDSYCSDLVI 400
 Db 398 D-PLELEYLREKKNKVDVTPKSDPTKSVQIPKVPYPNGIVYPLPLTDIHNSLAAD--- 453
 QY 401 NQIKNRY-KILNDNLNPSINEGTPTNTMKIPSDKLASIHEDNMFMFKITNYLKVGF 459
 Db 454 -NDKNSYGLMNPDPTEKINE-----KIITD-----NKRKIFT-----NNIK- 490
 QY 460 PDVRSTINLSGPGVVTGAYQDLLMFKDNSTWHLLEPELRFNPFEPKTKISQLETSITSL 519
 Db 491 -----KQIDLEEKINHTEQNKKLL 511
 QY 520 WSNQARAKSQFEYKKGYPFEGALGEDNLDLFAQVTLDKDYVSKILSSMKTRNKE--- 576
 Db 512 EDYB--KSKDYELLEKFYE-----MKFNNN--FDKDVVD-KIFSARYTYNVEKOR 558

Matches	350;	Conservative	287;	Mismatches	613;	Indels	533;	Gaps	88;
Qy	36	NMSESSVVEKYKLKADINNDNYINTYTKSGRKNALKKFKXYLTMEVLEKNNLSLPVE	95						
Db	106	NPSDNSSDNKTYADLAKHRVQNYLFT-----IKELKYPFLDITNHWLT-LS	152						
Qy	96	KNLHFIWIGGQINDTAINYINQWQVNS-DYTVKPVYDSNAFLINTLKKTIIVSATTNLT	154						
Db	153	KNV-----DGFKYLIDGYEBEINELLYKLANFYD-----LLRAKLNDACANSYC	195						
Qy	155	E-SPRELNDPEFDYK-----FYKRMELIYDKQHPIDYKYSQISENPEFIIDNIKTY	209						
Db	196	QIPNLKTRANELDVLKIVGYRKPLDNIKDNVGMEDYIKGN-----KTT	242						
Qy	210	LSNEYSKOLEALNKYIEESLNKITANNNDIRNLEK--FADEDLRLVLYNOELVERWNLAA	267						
Db	243	IAN-----INLEIGSKTIDONKNADNEEGKKLYQAQYNLFYINQKLOEAHNLIS	294						
Qy	268	ASDILRISMLKEDGGVYLDVILPGIQDPLFKSI--NKPDSITNTSMEMIKLEAIMKYKE	322						
Db	295	VLE-KRIDTLKKNENIKKLIEDIKITDAENPTTGSKNPLPENK-----KK	341						
Qy	326	YIPGYTSK-----NFDMLDEEVSQSFESALSXSKSDSEIFPLDDIKVSPLEVKI	375						
Db	342	EVGHEEKIKSIKTIKFNIDSLFTD-PLELEYLYLREKNKVDVTPKSDPTKSQVLPKV	400						
Qy	376	AFANNSVINQALISLSDSYCDLVINQIKNRY-KILNDLNPSINEGTFNTTKWIFSDK	434						
Db	401	PYPNGIVYPLPTDIHNSLAD---NDKNSYGDLMHPDTKEKINE-----KIYD-	447						
Qy	435	LASISNEDNMFMKITYNLKVGFPADVRSITINSGPGVYTGAYQDILLMFKONSTNIHL	494						
Db	448	-----NKRKIFPI-----NNIKKQIDL-----	464						
Qy	495	EPELRNPEPPKTKISQLTEQITSLSWSPNQARAKSQPEEYKKGYPFEGALGEDNDLDPQN	554						
Db	465	---BEKNINHTKQBNKKGLE-----DYEKSKQYEELEKPYE-----MKFNNN	505						
Qy	555	TVLDKDYVSKKILSSMKTNRKE---YIH-----YIVQLOGDKIYSBASCNLFSKDP	602						
Db	506	--FKQDVD-KIFPARYTYNVEKQRYNNKSSNSNSVYNQVKKALSY-----LED	554						
Qy	603	YSSILYQKNIEGSTAYYYY---ADAEIKEDKIRIPYOISNKRN--IKLFTIG--HGK	655						
Db	555	YS---LRKGISEKDFNYHTLTKTGLEADIKLTE---BIKSENKILEKNFKGLTHSA	606						
Qy	656	SEFNWTFDANLDV-DSLSSEIETILNLAKADISPKYIEINI-----LCGNMFSYSIYABET	710						
Db	607	N-----ASLEVSDDIVKLOQVKVLLIKKIEDLRK-IEFLKNAQLKDSIHVPNIYKPN	658						
Qy	711	YPGK---LLLKIKDRVSELMPISIQ-----DSIT---VSANQVEVRIINEEG--KR	752						
Db	659	KPEPYIYLVKVEDVKKEPIPKVKMDLKEQAVLSSITQPLVAASE-----TTDEGGHST	714						
Qy	753	EILDHSGKWINKEESIIND-----ISSKEYISFNPKENKIIVAKSKYLHELSTL	800						
Db	715	HTLSQSGETEVEETVEETVGHITTTVTITLPPKEESA--PREVKVENS-----	763						
Qy	801	LOEIRNNANSSDIDLEKKVMTECEINVASN-----IDRQIVEGRIBEAKNLTSDSI	852						
Db	764	---IEHKSNDNSQALTTVTYLLKLDDEFITKSYICHKYLVSNSMDQKLEVLNLTPEEE	820						
Qy	853	NYIKNEFKL-----IESISLSYDLKHQGLDSDHFIISFEDISKETENGPRIRFINKETGN	907						
Db	821	NELASCDPLDLLFNQNNIPAWSYLSYDSMMIDLOHL--PFELQYKEMYIYHLKLEENHI	878						
Qy	908	SIFIETEKEI-----FSEYATHIS--KEISNIKDTIPDN-----VNGKLVK	946						
Db	879	KCLLEQKQITGTSSTSGNVTWNTAQSAHNSNQNSQNASSTNTQNGVAVSSGPAVV	938						
Qy	947	KVNLDAAEHVNTLNSAPFIOSLIEY-NTTK-ESLSNLVSAMKVQVYQAQLFSTGLNTITD-	1003						
Db	939	EESHDPITVLSISNDLKGIVSLNMLGNKTKVPNPLTISTTMEKFFYENILKKNNDTVFND	998						

RESULT 6	MSPl_PLA	MSP1_MSP
ID	AC	P19
DT	DT	01-1
DT	DT	01-1
DT	DT	01-1
DE	DE	Mer
DE	DE	(PM)
GN	MSP	MSP
OS	Plac	Plac
OC	OC	Euk
OX	OX	NCB
RN	RN	[1]
RP	RP	SEQ
RA	RA	Cer
RT	RT	"A
RL	RL	pre
RN	RN	(2)
RP	RP	SEQ
RX	RX	MED

Db 1330 FTNRLND---ILNRLKRRYFLDVLSDLMQFKHSSNYIED-----S 1374
 QY 1425 KYLLSGCMKLIENSSDIQKIDH-IGFNGHQY---IFYSYIDNETHYNGFI----- 1475
 Db 1375 FKLLNSEQNTLLKYYIKESVENDIKPAQGISYEVKVLAKYKDDLESIKVKEKE 1434
 QY 1476 -----DYSKKEGLFTAESNESIIRNIYMPDSNNFIYSSKDLKDIRIIN 1520
 Db 1435 RPSSPTTPPSAKTDEQKESKFLPFLTN---IETLY-----NNL-----VN 1474
 QY 1521 KDVKLLIGNYFKDMKVSFTIEDTNTIK 1551
 Db 1475 K-----IDYLYNLKAKINDCNVEK 1494

RESULT 7

MSPI_PLAFM STANDARD; PRT; 1701 AA.
 AC P08569;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMMSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxId=70153;
 RN [1]
 RN [2]
 RN [3]
 RA Tanabe K.; MacKay M.; Goman M.; Scaife J.G.;
 RA "Allelic dimorphism in a surface antigen gene of the malaria parasite
 RT Plasmodium falciparum.";
 RL J. Mol. Biol. 195:273-287(1987).
 RN [2]
 RP REVISIONS TO 1403; 1569 AND 1629.
 RA Tanabe K.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-115 FROM N.A.
 RX MEDLINE=86136024; PubMed=3004972;
 RA MacKay M.; Goman M.; Bone N.; Hyde J.E.; Scaife J., Certa U.,
 RA Stunnenberg H., Bujard H.;
 RT "Polymorphism of the precursor for the major surface antigens of
 RT Plasmodium falciparum merozoites: studies at the genetic level.";
 RL EMBO J. 4:3823-3829(1985).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; X05624; CAA29112.1; --
 DR InterPro; IPR006209; EGF_like.
 DR Pfam; PF00008; EGF; 1.
 KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL
 FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AP96EA98 CRC64;
 Query Match 3.6%; Score 313.5; DB 1; Length 1701;
 Best Local Similarity 19.9%; Pred. No. 8.6e-05;
 Matches 357; Conservative 284; Mismatches 601; Indels 549; Gaps 91;
 QY 36 NMSESSVVEKYKLDKIDNNLTNTYKSGSRNALKKFKKEYLMEVLEKNNLSLTPVE 95
 Db 106 NPSNDSSTNTYADLRHVQNYLFT-----IKELKYPEFLDLNRLMLT-IS 152
 QY 96 KNLHFIWIGQINDTAINYNQKDVNS-DYTVKFVYDSNAFLINTLKKTIIVESATNTL 154
 Db 153 KNV-----DGFKYLIDGYEBEINELLYKLNFFYD-----LLRAKLDACANSVC 195
 QY 155 E-SFRENLDPEFDYK----FYKRMEIIVDKQHFIDYKQSTEENPEFIIDNIITY 209
 Db 196 QIPFNLRANELDVLKIVFGYRKPLDNKDVGMEDYIKKN-----KTT 242
 QY 210 LSNEYSKDLKALNKYIEESLNKITTANNNDIRNLEK--FADEDLVRLYNQELVERNLAA 267
 Db 243 IAN-----INELIEGSKTTIDQNKADNEEGKKLYQAYNLFIYKQLOEHLNIS 294
 QY 268 ASDILRISMLKEDGGVYLDVILPOIQDPDFKSI--NKPDSITNTSWEMIKLEAMKVK 325
 Db 295 VLE-KRIDTLKKNENIKKLEDDIKITDAENPTTGSFKNPLPENK-----K 341
 QY 326 YIPGYTSK-----NFDMLDEVOVSFESALSCKSEIFLPLDDIKVSPLEVKI 375
 Db 342 EVEGHEEKIKETAKTIKENIDSLFTD-PLELEYLRKKNKKVDVTPKSDPTKSVQIPKV 400
 QY 376 AFANNSVINQALISLKSVCSDLVINOIKNRY-KITNDNLNPSINEGDTFNTMTIFSPK 434
 Db 401 PYPNGIVYPLPLTDIHNSLAAD-----NDKNSYGDLMNPDTKEKINE-----KIITD- 447
 QY 435 LASISNEDNMFMFIKITNLYKVGFPADVRSTINISGPGVYTGAYQDMLLMKDNSTNIHLL 494
 Db 448 -----NKKERKIFI-----NNIKQIDL----- 464
 QY 495 EPELNFPPKTKISQLTQEITSLSWQNAQAKSQFBEYKGYFEGALGEDNDLDFQDN 554
 Db 465 --EKNINHTKEQNKKLLE-----DYEKSKDYBELLEKPYE-----MKFN 505
 QY 555 TVLDKDYYSKILSSMKTRNKE---YIH-----YIVQLODKISVEASCNLFSPKDP 602
 Db 506 --FDKVDVY-KIFSARYTNVEKORYNNKFSNNNSVYNVQKLKALSY-----LBD 554
 QY 603 YSSILYQKNIEGSEYAYVYV--ADAEIKEDIKYRIPVOISNKN--IKLTFIG--HCK 655
 Db 555 YS--LRGIGSEKDFNHYTTITKLEADIKKLE-----EIKSSENKLEKKNFKGLTHSA 606
 QY 656 SEFNTDTFANLDV-DLSSEIETILNLAADISPKYIEINL----LGCNMFYSYSIYAET 710
 Db 607 N-----ASLEVSIVDKVQVQVLLIKKIEDLRK-IELFLKNAQLKDSIHVNYKPN 658
 QY 711 YPGK-----LLKIKDRVSELMPISQ-----DSIT---VSANQVEVRINEEG--KR 752
 Db 659 KPEPYLLVLKKEVDKLKEFIPKVKMDLKEQAVLSSITQPLVAASE-----TTEDGGHST 714
 QY 753 EILDHSGKWINKEESI IKD-----ISSEKYSFNPENKNIIVKSKYLHELSTL 800
 Db 715 HTLSQSGETEVEETEVEETVGTHTTTITLPPKEESA--PKEVKVENS----- 763
 QY 801 LOEIRNANSSDIDLEKKVMLTECE-----INVA-SNIDRQIVEG---RIEBA 844


```
Db 764 ---IEHKSNDNSQALTKTVLKKLDFLTKSYCHYILVSNSSMDQKLEVLNLTPEE 820
Qy 845 KNLTS---DSINVIKNEFKLIESISLSL-YDLKHQGLDSDSHFISFEDISXTENGFRIR 899
Db 821 KELKSCDPLDLLFNQNNIPAMVSLYDSMNLDQH-----LFFELYKEMIIYU 870
Qy 900 FINKETGNSIFITEKEI-----FSEYATHIS--KEISNIKOTIFDN--- 939
Db 871 KLKEENHIKKLBEQKQITGTSSTSPGNTTNTVAQSAATHSNQCSNASSTINTQNGVA 930
Qy 940 -VNGKLVKVNDAEAVNTLSAPFIQSLIEV-NTTK-ESLSNLSVAMKVQVYQLPFT 996
Db 931 VSSGPAVVEESHDPVLVSLNDLKGIVSLNKGKTKVPNPLTITSTEMEFYENILKN 990
Qy 997 GLNTITD-----ASKVVE-LVSTALDETIDLLPTLSEGLPIIATIDGVSL----- 1041
Db 991 NTDFYNDIDIKQFVKSNSKVITGLTETQKVALNDEIKKLDKDTQLSPDLNKNYKCLKLDRUF 1050
Qy 1042 -----GAAIKELSETNDPLLRQIEIAKIGIMAVNLTAATAIVTSALGIASGFSIL 1092
Db 1051 NKKKELGQDKWQIKKLT-----LLKEQLESKLSL-----NNPHNVLQNFSVF 1093
Qy 1093 L--VPLAGISAGIPSLVNNELILOKATKVIDYFKHISIAETGAPTLLDDKIKMPQDDL 1150
Db 1094 FNKKKEAEIAETENTLENTLKILKHVKYGLVKNYNGESSPLKT-----LSEVSIQEDNY 1147
Qy 1151 -----VLSEIDFN-NNSITLKGCEIWRAGSGGHTLDDIDHPSSPSITYRKPWLSI 1202
Db 1148 ANLEKPRALSKIDGKLNHLGKGLKLSLSSGLHLITE----- 1187
Qy 1203 YDVLNLIKKEKIDFSDKDLVLPNAPNRFVGYEMGTWPGFRSLNDGDKLLDRIDHYEQF 1262
Db 1188 -----LKEVI---KNKNYTGNSP-----SENKKVNEALKSVEN--- 1218
Qy 1263 YWRYFAPADALITKL-----KPRYEDTNVRINLDGNTSRFIVPVITTEQIRKNLSVSYFG 1318
Db 1219 -----FFPEAKVTTVVTPQPDVTPSPLSVRVSGSGS-----TKERTQIPTSGSLI- 1265
Qy 1319 SGGSYLSLSPYNNMIDNLV-----ENDTW-----IDVDNVVKNITIEDS 1360
Db 1266 TELQVQVQLQNTDEEDDSIVLPIFOESDNDDEYLDQVVTGSAISVTMDNIUSGFENEYD 1325
Qy 1361 EIQ-----KGELIENLS-KLNTIEDNKILNNH--TINFYGDINESN---RPI 1402
Db 1326 VIYKPLAGVYRSLKQIEKNIITFNLDN---ILNSRLKRRKYFLDVLESDLMQFKHI 1382
Qy 1403 SLTFSILEDINIIEIDLVSQYKILLSCNCKLNSSDIOQKIDH-IGFNGEHQKY-- 1459
Db 1383 SSNEYIIE-----SFKLNSRQKNTLLKSYKIKESVENDIKFAQEGISYVE 1430
Qy 1460 -IFYSYIDNETKNGFI-----DYSKKEGLFTAEPFNSIIRNI 1497
Db 1431 KVLAKYKDDLESIKKVIKEKEKFPSPPTTPPSAKTDQEKESKFLPFLTN---IETL 1487
Qy 1498 YMPDSNNLIYSSKDLIRINKGD-----VKLLIGNYFKDDMKVSLSTIEDTNTI 1550
Db 1488 Y-----NNL-----VNKIDDYILNLKAKINDCNVEKDEAHVKIT-KLSDLKAI 1529
Qy 1551 --KLNGVYLDNGVAQILKFMNAKSALNTSLSNMFLESINIKNIFYNNL 1599
Db 1530 DDKID-LFKNTNDFAIKKLIND-----DTKKDMLGKLLSTGLVQNFNPNTI 1574

RESULT 8
MSPI_PLAPP STANDARD; PRT; 1726 AA.
AC P50495;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (GP195).
GN MSP-1.
```

```
OS Plasmodium falciparum (isolate Palo Alto / Uganda).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57270;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89005525; PubMed=3049134;
RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA Siddiqui W.A.;
RT "Plasmodium falciparum: gene structure and hydropathy profile of the
RT major merozoite surface antigen (gp195) of the Uganda-Palo Alto
RT isolate.";
RL Exp. Parasitol. 67:1-11(1988).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M37213; AAA29611.1; -
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196174 MW; 5859CEBF9A026 CRC64;

Query Match 3.6%; Score 312.5; DB 1; Length 1726;
Best Local Similarity 19.9%; Pred. No. 9.6e-05;
Matches 350; Conservative 286; Mismatches 600; Indels 523; Gaps 90;

Qy 51 DINNLTNDVNTYKSGRNKALKKFKKYLTWVLELKNLSLTPVEKNLHFTWIGQINDT 110
Db 154 DLKHRVRYLFT-----IKELYPFLDTLTHMLTCLD-NIH----- 189
Qy 111 AINY-INQWQVNSDYTVKPYVDSNAFLINTLTKTIVESATNTLE-SFRENLDNPEFDY 168
Db 190 GFKYLIDGYEBIN-----ELLYKLN-FYFDLRAKLANDVCANDYCCQIPFNKIRANELD 243
Qy 169 NK----FYRKMEIITDKQKHFIDYKSOIEENPEFIIDNIKTVLSNEYSKDLKALNKY 224
Db 244 LKLVFGYRKPDLNKKVNGWEDYIKKN-----KTTIAN-----INEL 282
Qy 225 IEESLNKITTANGNDIRNLEK---FADEDLVRLYNQELVERWNLAASDIIRISMLKEDG 281
Db 283 IEGSKKTIQKNADNBEKGLKLYAQYDL-SYNKQLEEAHNLISVLE-KRIDTLKNE 340
Qy 282 GVLVDVLDLPQIQ-PDLFKSINKPDSITNTSWEMIKLEAINMKYKYEIPGYTSKQFMDLDE 340
Db 341 NIKELDKDINEIKNPPPPANSNGTNTLDDKNKKEEHEE--KIKE-TAKTIKFNIDSFT 397
Qy 341 EVQRFSEFALSCKSKDSEIFLPLDDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVI 400
```

Db 398 D-PLBLEYLRKKNKVDVTPKSDQPTKSVQIPKVPYNGIYVPLPLTDIHNLSAAD--- 453
Qy 401 NOIKNRY-KILNDNLNPSINEGDTFTMTKIFSDKLASINEDNMWMLKINTLYKVGFA 459
Db 454 -NDKNSYGLMMPDTEKINE-----KIITD-----NKRKIFI-----NNIKF----- 490
Qy 460 PDVSTRINLSGPGVYTGAYQDILMFKDNSTNIHLLPELRNPEFPKTKISQUTEQITSL 519
Db 491 -----YIVQLQDKITSYEASCNLFKDPYSSILYOKNIEGSETAYVVYV---A 624
Qy 520 WFNQARAKSQPEEYKGYFEGALGEDNDLPAQNTVDKOVSVKILSSMKTRNKE--- 576
Db 512 EDYE--KSKOYEELELFYE-----MKFNNN--FDKQVVD-KIFSARYTYNVEKOR 558
Qy 577 YIH-----YIVQLQDKITSYEASCNLFKDPYSSILYOKNIEGSETAYVVYV---A 624
Db 559 YNKFSSNNNSVNVQKUKALSY-----LEDYS-----LRGISEKDFNHYITKLTGL 607
Qy 625 DAEIKEDKIRIPYOISNKR--IKLFTIG--HGKSEFNTDTFANLDV-DSLSSBIETIL 679
Db 608 EADIKKLT-----EIKSSENKILEKNPKGLTHSAN-----ASLEVYDIVKLQVKVL 655
Qy 680 NIADADISPKYTEINL-----LCNMFYSIYAETYPKG-----LLKIKDORVSELMV--- 728
Db 656 LIKKIEDLRK-IEFLKNAQLKXDSIHVPNIYKPNKPPYIYLVKKEVDKLEKPIPKVK 714
Qy 729 -----SISODSITVSANOYE-----VRINEEGKREILDHSGKWINKESIIKDI 772
Db 715 DMLKEQAVLSITPLVAASETTEDGGHSTHTLSQGETEVTEETEE---TEETV--GH 769
Qy 773 SSKEYISFNPKENKIIVKSKYLHELSTLLOEIRNNANSDDILEKKVMLTECEINVAASN- 831
Db 770 TTTVTITLPPKEVKVENS-----IEHKSNDNSQALTKTVVLKLDLFELTKSY 817
Qy 832 -----IDRQIVEGRIBEAKNLTSDSINYIK-----NEFKLIESISLSL-YD 871
Db 818 ICHKIYILVNSMSMDQKLELVNLTPEEENELKSCDPLDLFFNIQNNIPAMVSLYDSMND 877
Qy 872 LKHQGLDDSHFISPEDISKTEGFRIRFINKETGSIPIETEKEI----- 917
Db 878 LQH-----LFFLYQKEMIIYLLHKLKEENHIKKLEEQKQITGTSTSSPGNTTV 927
Qy 918 -FSEYATHIS--KEISNIKDTIFDN-----VNGKLVKVNLDAAHVNTLNSAFFTOSLIE 970
Db 928 NTAQSAHNSNQNSASNTTQNGVAVSSGPAVVEESHDPDLTVLSISNDLKGIVSLN 987
Qy 971 Y-NTTK-BLSNLSVAMQVQVYAQLFSTGLNTITD-----ASKVVE-LVSTALDETI 1019
Db 988 LGNKTQVNPPLTISTTEMEKFYENILKNDTYFNDIDIKQFVKSNSKVITGLTQKNALN 1047
Qy 1020 DLLPTLSEGLPIIATIIDGVSL-----GAAIKELSETNDPILRQIEAKI 1064
Db 1048 DEIKKLDKTLQSLFQYLNKYKLDRLFNKKELGQDKWQIKKLT-----LLKEQLESKL 1102
Qy 1065 GIMAVNLTAASPAIVTSALGIASGFSILL--VPLAGISAGISPLVNNELILODKATKVID 1122
Db 1103 NSL-----NNPNVLFQNSVFPNKKKEAIEAETENTLENTKILLKHVYGLVY 1150
Qy 1123 YPKHISLAETEGAFLLDDKIIMPODDL-----VLSEIDFN-NNSTLKGKCEIWRABG 1174
Db 1151 YNGESSPLKT-----LSEVSQIOTEDNVANLEKFRVLSKIDGLNDLHLGKGL----- 1199
Qy 1175 GSGHTLTDHDFPSSPSITKPKWLSIYDVNLNIKEKIDFSKDLMLVLPNAPNRVFGVEM 1234
Db 1200 -----SFLSS-----GLHQLITELKEVI---KNQNYTGNSP----- 1227
Qy 1235 GWTGFRSLDNDGTKLDIRDHYEGQFVWRYFAFIADALITKL-----KPRYEDTNVRIN 1290
Db 1228 -----SENKKVNEALKSYEN-----FLPEAKVTVTVPQDPDVPPLSVR 1269
Qy 1291 LGNTRSFIVPITTEQIRKKNLSYFGSGGSYSLSLSPYNNMINDNLV-----END 1342

Db 1270 VSGSGS-----TKEETQIPTSGSL--TELQVVVQLQYDEEDSLVVLPIFGSESDND 1322
Qy 1343 TWV-----IDVDNVKNITIESDEIQ-----KGELIENILS-KLINEDN 1380
Db 1323 EYLDQVVTGEAISVTMDNILSGFENEYDVIVLKPLAGVYRSLLKQIEKNIPTFNUNLND- 1381
Qy 1381 KIILNNH--TINFYGDINESN-----RFTSLTFSILEDIINIIIEIDLVSYSKILLSGNOM 1434
Db 1382 --ILNSRLKRYFLDVLDESLLMQPKHSSNEYIIE-----SPKLLNSEQKN 1427
Qy 1435 KLIENSIDIQKIDH-IGFNGEHQY-----IFSYIDNETKNGFI----- 1475
Db 1428 TLLKSYKIKVESVENDIKFAQEGISYEBVLAKYKDDLESIKVKIKBEKKEFPSSPTTP 1487
Qy 1476 -----DYSKKEGLFTABFSNESIIRNIYMPDSNNFIYSSKDLKDIRINKGD----- 1523
Db 1488 PSPAKTDDQKESKELPFLTN---JETLY---NNL-----VNKIDDYLN 1527
Qy 1524 -VKLLIGNYFKDDMKVLSFTIEDNTI--KLVGYLDENGVAQILKFMNNAKSALNTSN 1580
Db 1528 KAKINDCNVKEDEAHVKIT-KLSDLKADDDKID-LFKNHNDFDAIKKLIND-----DTCK 1580
Qy 1581 SLMPFLESINIKNIFYNL 1599
Db 1581 DMLGKLLSTGLVQNPNTI 1599

RESULT 9
MSPI_PLAFW
ID MSPI_PLAFW STANDARD; PRT; 1639 AA.
AC P04933;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMNSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5848;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86014355; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites."
RL Nature 317:270-273(1985).
RN [2]
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (-!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X02919; CAA26676.1; -.
DR PIR; A24594; A24594.
DR PIR; S05603; S05603.
DR PDB; 1CEJ; 28-MAY-99.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.

KW	Malaria; Merozoite; Polypotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor; 3D-structure.	
KW	SIGNAL	1 19
FT	CHAIN	20 1639
FT	CARBOHYD	116 1116
FT	CARBOHYD	268 268
FT	CARBOHYD	764 764
FT	CARBOHYD	768 768
FT	CARBOHYD	783 783
FT	CARBOHYD	844 844
FT	CARBOHYD	920 920
FT	CARBOHYD	964 964
FT	CARBOHYD	1058 1058
FT	CARBOHYD	1165 1165
FT	CARBOHYD	1174 1174
FT	CARBOHYD	1445 1445
FT	CARBOHYD	1526 1526
FT	CARBOHYD	1539 1539
FT	SEQUENCE	1639 AA; 187618 MW; 2C255B8616C87F68 CRC64;
Query Match	3.6%; Score 309; DB 1; Length 1639;	
Best Local Similarity	19.5%; Pred. No. 0.00012;	
Matches 351; Conservative 280; Mismatches 554; Indels 618; Gaps 94;		
QY	36 NMSESSVVEKYKLKDINNLNTYNTLYTKYSGRNKALKKFKXYITMVEVLELKNLSLPVE	95
DB	112 NPSDSSDSDAKSYADLKHVRNYLLT-----IKELKYPQLFDLTHMLTCLD	159
QY	96 KNLHFIWIGQINDTAINY-INQKDVNSDYTVKVDNSAFNLTLTKKTVIVESATNNTL	154
DB	160 -NIH-----GFKYLDIGYEIN-----ELLYKLN-FYDPLRAKLDNCANDYC	201
QY	155 E-SFRENLDPPFDYNNK-----PYRRMBIIYDKQGHFDIYKYQIENPEFIIDNIITKY	209
DB	202 QIPFNKLIRANELDLVKLVFGYRKLNDIKDNGVMEDYKKN-----	245
QY	210 LSNEYSKDLKALNKYIEESLNKITANNGNDIRNLEK-----FADBDLVRLNQBELVERNL	265
DB	246 -----KKTIIENELIEES-KKTIDKNKATKEEKKLYQAQYDL-SIYNKQLEEAHNL	298
QY	266 AASDILRLSMLEKGGVLDVILPGIQ-PDLFKSINKPSIINTSWEMIKLEAIMKYK	324
DB	299 ISYLE-KRIDTLKKNENIKELDLKNEIKNPPANSNGTNTL-----LDKNKKIE	348
QY	325 EYIPGYTSKNFMDLDEVQSPESALSCKSDKSEIFLPLDDIKVSPLEVK--IAPANNSV	382
DB	349 EH-----EKEL-----KETAKTIKFNIDSLFTDPLELEYLREKQKNI	386
QY	383 INQALISLKDSCDLVINOIKR--RYKILNDNLNPSINEGTDFTNTTKIFSPDKLASIN	440
DB	387 DISAKVETKES---TEPNEYENGVTYPLSYNDINNALNELNSFG-----	427
QY	441 EDNMFMWIKITVYLVKGFAPVRSINISGPGVYTGAYQDILLMPKDNSTNIHLEPELRN	500
DB	428 -----DLNPFDTY-----KPSKNIY-TDNERKK	451
QY	501 FEPPKTKISOLTEQ-EITSLWSFNQARAKSOPFEEYKGYFEGALGEDDNLDPQNTVLDK	559
DB	452 F-----INEIEKIKI-----EKKKIESDKSYEDRSKSLND-----ITK	486
QY	560 DYVSKKILSSM---KTRNKEYIHYIVQLQGDKISYEAASCNLFPSKDPYSSILYQK-NIEGS	615
DB	487 EY-EKLLNEIYDSKFNNDILDTNFEKWMGKRYKYVKE-KLTHNTTASVENSKHNL--	541
QY	616 ETAYYYYVADAIEIKDKYRIPIYQISKNENIKLTFIGHGKSEFNTDTFANLDVDSLSSEI	675
DB	542 -----KLTKALKYMEDYSU-----RNIVV-----EKLKYYKQNL-ISKIENEI	578
QY	676 ETILNLAKADISPKYTEINLLGCONFYSIYAEBETYPGKLLLKIKDRVSELMPSISQDSI	735
DB	579 ETILVENIKDEE-----QLFEKKITKDNKPKDEKILEVSDIV-----	615
QY	736 TVSANQYVREINBEGKREILDHSGKWINKESIIKDISKEYI---SFPNPKEN-----I	787

```

RESULT 10
YCF2_MARPO
ID YCF2_MARPO STANDARD; PRT; 2136 AA.
AC P09375;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN YCF2.
OS Marchantia polymorpha (Liverwort).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiaceae;
OC Marchantiaceae; Marchantia.
OC NCBI_TaxID=3197;
OX [1]
RN
RP SEQUENCE FROM N.A.
RA Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA Ozeki H.;
RT "Chloroplast gene organization deduced from complete sequence of
RT liverwort Marchantia polymorpha chloroplast DNA.";
RL Nature 322:572-574 (1986).
CC -!- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; X0465; CAA28078.1; -
CC PIR; S01591; A05037.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC Pfam; PF00004; AAA_1.
CC SMART; SM00382; AAA; 1.
CC Chloroplast; Hypothetical protein.
KW SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;
SQ
Query Match 3.5%; Score 306.5; DB 1; Length 2136;
Best Local Similarity 20.3%; Pred. No. 0.00021;
Matches 387; Conservative 271; Mismatches 629; Indels 623; Gaps 102;
QY 27 ILNALDEEYHNMSESSVEKYLKLDINNL-TDNYL-----NTYKSGRNLKALKKPKLEYLT 80
DB 309 IFKNLQNP-NESDKKLIIESFFLLKIKGNLYFKNYIEFVTWQSYKKDCLD--FNKNELNN 365
QY 81 MEV-LELKNNSLTPVEKNLHFTWIGGQINDTAI-----NVINQWQVNS--DYTVKPVYDS 133
DB 366 SBIYKIEBELFSDIYKFSKYLIEGKSKTKTIKQSFNNNIYKKLNSFNFTPIYFDS 425
QY 134 NAFLINTLTKKTIVESTATNT--LESFRENLDPEPDY----NKFYKRMELIYDKKHPI 187
DB 426 NLLFDMLKKNY--INKNPLKSP-----LIYSSISNQF-----ILFFQKNSK 468
QY 188 DYYKQIEENPEFIIDNIITKLYS--NEYSKDL-----EALNKYTEESLNKIT 233
DB 469 SFNKNLVKNSKDVITNVPSKENKTEINNFPSKIYAFPEILSINEIDNKFV---INKIS 525
QY 234 ANGNNDIRN-----LEKPADELDVRLYNQELVERWN-----LAAASDILRIS-ML 277
DB 526 LKNINKKKQKRYLNKIKSSDNFRFINLWKIKNYSQVFNNSFLNPAFELQOYYL 585
QY 278 KEDGGVYLDVDLPGIQDPLFKSINKPDS-----ITNTSEMIKLEAIMK----- 322
DB 586 KGNILF-----FKKJNEVSNFFYQYKCKKLNIFLKASLEKILKCRNK 632
QY 323 -----YKEYIPGVTSKNFDMLEBQVRSFESALSSKSDKSEIFPLDDIKV----- 368

```

```

Db 633 KFTISIKLKFYKKNLNGEYKTESQILQNEKELNKKRKNQFNP--NIKILSFYNS 690
QY 369 SPLEVKIA---PANNVINQALISLKDSYCSDLVINQIKNRKYKILNDNLNPSINBGTDFN 425
Db 691 SKKNIYLQNKYFFNKNLNNKLITWK-----KISNKLIVSNSEYKNKIWN---K 736
QY 426 TTMKIFSDKLASISNEDNMFMKITYLNKLVGFADPVRSTINLSGPGVVYTGAYQDL---- 481
Db 737 KNMKFFSFSKSVL--DTFFFNKKSFNITVIF--DKLKIKQLN-----FQEIQKIL 784
QY 482 ----LMPKDNSTNIHLLSPELNFPFPKTKI---SOLTEQEITSLWSFNQARAKSQFBEY 534
Db 785 NCFSLFF--NSKNI-----KTKIPKNSYFINENLTTTFSFND----KEFNIF 826
QY 535 KKGYPEGALGEDNDLDAQNTVLDKDYVYSKTLSS-----MKTRNKVEIHYIV 582
Db 827 FLELFISINNDFLMRFKCYLYRYIKDBELTFNPIENRQLQLNFFFEKTKILTIDF-- 884
QY 583 QLQGGKISYEASCNLFSDPYSSILYQKNISETAYY----YYVADA-----EIKV- 630
Db 885 -LQDEPANY-----NNRFIFHEKTKIKNNLLYLRLKIFLKDKNFLLINEIKSF 935
QY 631 IDKYR---IPQISNKRNIKLTFIGHGKSEFNTDTPANLDDVDSLSEIEITILNLAkadIS 687
Db 936 ICKNNLFKISQSLNVLVKSNSY-----KPPDNIFNFHFLKQKKNIEILN-----N 983
QY 688 PKYIINLL-----GCNMFYSIYABET-----YPGKL 715
Db 984 QNYFEKSLKTKTYLKNLNNNSYKFSYKIFIFQLNLNKNKYKTFOWISELIFYSKNL 1043
QY 716 LKIKDRVSELSPISQSDITVSANOYEV--INEEGKREILDHSGKWINK----- 764
Db 1044 NYKIQNKIEKNYCNKN---ISYKKKIKITVPEKKNLFQTNNSWPFLEWEVNTYI 1100
QY 765 -----ESSIIDISSKISYISNPKNKIIVKSKYLHELSTLQEIIRNNANSIDLEKKV 819
Db 1101 LIQIIQETFFQITDVLVFP-----KKKIIIEKN-----LKFFLK-----SKKI 1138
QY 820 MLTEC-----EINVASNIDRQIVEGRIEAEKNLTSDSINVIKNEPKLIES 864
Db 1139 SLKTLSFHNFKLKWNLRFPNEINYYKNY-----LNLFLMSDFNLINN 1180
QY 865 ISDSL-----DLKHQN-----GLDSH--FISFEDISKTENGPRIFRINKET-G 906
Db 1181 -CNNLYWIFSLVIFLYQKIFSIILIGSDCFHLWKWPEIIQYLTDRSRSLYFKLTR 1239
QY 907 NSFIETEKEIPSEYATHISKEISNIKOTIDPNVNGK--LVKKVNLDAHAHEVNTLNSAPP 964
Db 1240 NKTALNKTENLLSYFQNLTHYITNKFYLLTKGNLKNKTLNKTLDLSRRKRKL----L 1295
QY 965 IOSLI-----EYNTKE-----SLSNLSVAMKVQVYLAQLFSTGL--NTITDA 1004
Db 1296 VOSLTIHKNIQNYGFELNSKQFTSYFYQYQITNOQGLLYFOYLAQFPQKNLNNSLDLA 1355
QY 1005 SKVBE-----LVSTALDET--IDL-----LP-----TLSEGLPIATIIDGVSLGA 1044
Db 1356 NKWIVSFHWKIFSSQKLRQTNWIELGQNIYVPVQFGLSYSGKILLIGPIETGSY--L 1413
QY 1045 IKEL-SETNDPLRLQREIEAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGISAGI 1103
Db 1414 IKNLAAESVPLFKISIN-----KLLYNKPDVITE-----SWNNILIESLRL-- 1456
QY 1104 PSLVNNELIQDKATKVIDYFKHI-----SLAETEGATLLDDKLIIMPDDDLVLS- 1153
Db 1457 ----NLTLDFAKMSPCIWIQNIHQNLNVRNLTQVNESDPTFL-----GILLKY 1502
QY 1154 -EIDFN-----NNSITLKGCEIWRASGGSGHTLTDIDHFFSSP-----SITYR 1196
Db 1503 FOTDFSKTKKNIIIV-----GSTH-LPKKVDPALISPNRDLKIINVRLEFNISOR 1551
QY 1197 KPWLSIYDVNLNKEKIDFSDKMLVLPNAPNRVFGYEM-----GWTGFRSL 1243

```

Db 1552 KQFP-----LLKKKQFQKLENFLPNERGSRMTGYNLRDLSALTNEVLLISITKNRSP 1607
Qy 1244 DNDGTLKLRDRDHGQYRYFYAFIADALLTKLKPVEDTNVRINDGNTRSFIVPI 1303
Db 1608 DTDTLK-----AFHQIQFGLTY-----TNNKLNFD---RIFKIVY 1641
Qy 1304 TTEQIRKNLSYFGSGGYSLSLSPYNNIDLNVENDTWIDV---DNVKNITIED 1360
Db 1642 ---KVGKTIQNLIKSSWNIL-LNIGNFLWKNFYLSKWLEPSIDESIKELTILTH 1697
Qy 1361 EI-----QKGEIENILSKLNIEDNKIILNNHTINFYGDINESNRFISLTSILED 1411
Db 1698 ILACLAGTAARDSWFLKKAESL-LPIDKLVENDFT-----LAFSILES 1741
Qy 1412 INI-----IIEIDLVSYSYKIL---LSGNCKMLIENSDDIOQKDHGFGNGEHQYIPY 1462
Db 1742 FFEPPMLEICQTNVNSKKNKIIIEFTKNSNMIMQNGI-----FAIANKKFI-- 1789
Qy 1463 SYIDNETKNGF-----IDYSKEGLPTAFSFSNIESIIRNIYMPDSNNLFTYSSKOLKDIRI 1518
Db 1790 -YTONHLQYKSSLSQISFNKKQY---EFKQTS-----WSPRFWLSFRSNLFDWIKR 1840
Qy 1519 INKGDVKLAGNYFKDDMKVSLSTIEDTTIKLNGVYLDENGVAQILKPMNNAKSALNT 1578
Db 1841 PN-----DPEFSYKF-----GFTKKKEYLFSAN--LQK 1866
Qy 1579 SNSLMNLFESINIKNIFYNNLPDNI-----EFILDTNFIISG 1615
Db 1867 KNYGQFIEKKKEQLLYERILPRIRRRNVQELSESOFEBILLEEQFEILG 1916

RESULT 11
MSPI_PLAFK STANDARD; PRT; 1630 AA.
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Meroporphin of the precursor protein 1 precursor (Meroporphin surface antigens) (PMWSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate K1 / Thailand).
OX Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OC NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum meroporphins: studies at the genetic level";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RN REVISIONS, SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (potential).
CC -I- PTM: MEROPORPHIN SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa and 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROPORPHINS. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X03371; CAA27070.1; --
CC InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.

Malaria; Meroporphin; Polyprotein; Repeat; Signal; Glycoprotein;
Transmembrane; GPI-anchor.
KW SIGNAL 1 19 POTENTIAL.
KW CHAIN 20 1630 MEROPORPHIN SURFACE PROTEIN 1.
FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 755 755 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 774 774 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 835 835 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 955 955 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;
Query Match 3.5%; Score 306; DB 1; Length 1630;
Best Local Similarity 19.5%; Pred. No. 0.00016;
Matches 348; Conservative 278; Mismatches 544; Indels 618; Gaps 94;
Qy 51 DINNLTDNLTATYKSGRKNALKKPEKYLTVLELKNLSLTPVEKNLHFHWIGQINDT 110
Db 118 DLKHVRNLLT-----IKELYPQLFDTNHLTLCD-NIH----- 153
Qy 111 AINY-INQWQVNSDYTVKVFVDSNAFLINTLKKTIVESATNTLE-SFRENLDPEFDY 168
Db 154 GFYLLIDGYEIN-----ELLYKLN-FYDILLRAKLVNDVCANDYCOIFNLKIRNELDV 207
Qy 169 NK-----FYRKMEIYDKQKHFIDYKQIEENPEFIIDNIKTLYSNEYSKDLBALNKY 224
Db 208 LKKLVFGYRKPLDNKDNVGMEDYIKKN-----KKTIEINEL 246
Qy 225 TEESLNKLTANNNDIRNLEK-----FAEDLVRLYNQELVERWNLAAASDIILRSMKED 280
Db 247 TEES-KTTIDRNKNATKEEKKLYQAQYDL-SIYNQOLEEAHLNISVLE-KRITLKK 303
Qy 281 GGVYLDVILPGIQ-PDLFKSINKPDSITNTSWEMIKLEAMKYKEYIPGYTSKQNFMDLD 339
Db 304 ENIKELLDKINEIKNPPPPANGTENTL-----LDKNKKIEH-----E 342
Qy 340 BEVQSFASLSKSKSEIFLPLDDIKVSPLEVK--JAFANNSVINOALISLKSQSCSD 397
Db 343 KEI-----KEIAKTIFNIDSLFTDPLEYLYREKKNKIDISAKVETKES---- 388
Qy 398 LVINQIKN--RYKILNDNLNPSINGTDFNTMTKIFSDKLASISNEDNMFMFKITNYLK 455
Db 389 TEPNYPNGVTPYPLSYNDINNALNSLNSFG----- 418
Qy 456 VGFAPDVRSTINLSGPGVYTCAYQDLLMFKONSTNIHLLEPELRNFPFKTKISQUTEQ- 514
Db 419 -----DLINPFDYT-----KEFSKNLY-TDNERKKF-----INEIKKI 451
Qy 515 EITSLWSFNQARAKSQFEYKKGYPGALGEDNDLDFQNTVLDKDYVSKKLSSM---K 571
Db 452 KI-----EKKIESDKSKYEDRSKSLND-----ITKEY--EKLLEIYDSK 490
Qy 572 TENKEYIHYIVQLQDKTSYEASCNLFKDPYSSILYOK-NIEGSETAYYYYVADAIEKE 630
Db 491 FNNNIDLNTNFENQGMGRYSKYVE-KLTHHTTFASTENSKHNL-----KLTK 536
Qy 631 IDKIRIPQISKNRKNIKLTFIGHGKSEFNFTDTFANLDVDSLSEIETILNLAKADISPKY 690
Db 537 ALKYMEDYSL---RNIIV-----EKELKYKKNL-ISKIENIEITLVENIKOEE--- 581
Qy 691 IEINLLGNMFSYSYASSETYPGKLLLIKORVSELMPSISODSTVSNQANVEINERG 750
Db 582 -----QLFEKKITKIDENKPEKILEVSDIV-----KVQV----- 610

Qy	751	KREILDSHGKWINKESIIKDISEKEYI---SFPNKENK-----IIYKSKYILHELSTILLQ	802
Db	611	QVLLMMNKIDELKKTQILKNVELKHNHPVNSYKQENKQEPYVILIILKEEIKDKLVFMP	670
Qy	803	EIRNANSSDIDLEKKVMLTCEINVASNDRQIVEGRI-----BEAKNLTSDSI	852
Db	671	KVESLINE-----EKNIKTEGQ-----SDNSEPSTEGEITQOATTKPGQAGSALEGDSV	721
Qy	853	NYIKNEPKLIB-SISDSLYDLKHQ-----NGLDSDHFISSFEDISKTENGFRIPINK	903
Db	722	QAOAQEQOQAQPPVPVPEPAKAQVTPPAPVN-----NKTENVSKLDYLEK	768
Qy	904	--ETGNSIFTEETEKEIFSEVATHISKEISNIKOTIPDNVANGKLVKKVNLDAAEHVNTLNS	961
Db	769	LYEFANTSYL-CHKYILVSHST-----MNEKILK-----	796
Qy	962	AFFIQSLIEYNTTKESLNSISVAMKVQVYAQLFSTGLNTITDASKVVVELVSTALDETIDL	1021
Db	797	-----QYKITKEBESKLSSC-----DPLDL	816
Qy	1022	LPTLSEGLPIATIIDGV--SLGAAIKELSE-----TNDPLLROEIEAKIGIMA	1068
Db	817	LFTNQNIPWYSPFDSLNNLSOLFWEIYEKEMVCNLYKLKONDKITKNLESAK-----	871
Qy	1069	VNLTAASATIVTSALGIASGFSILLVPLAGISAGPSL-VNNELILOQKATKVIDYPKHI	1127
Db	872	-----KVSTSVKTLs-----SSSQPLSLUTPODKPEVSANDDTSHSTNNLSKLUPENI	920
Qy	1128	SLAETEGAP--TLDDKIIMPQDDVLVSEID-FNNNSITLKGCEIWRABEGSGHTLTDDI	1184
Db	921	LSLGNKNIIYOELLIGOKSSSENFYEKILKDSDTFYNESFT---NFVKSKADDINSLNDES	976
Qy	1185	DHPFSSPSITVRKPLWSI-YDVNLIKKEKID--PS-----KOLWVLP-----	1223
Db	977	KRKKLEBIDINKLTQLSPDLNLYKKLIERLPDKKTVGKYKMQIKKLTLLKEQLESK	1036
Qy	1224	-----NAPNRVF-----GYEMGWTGFRSLDNDGTKLLDRIDRHYEG-----	1260
Db	1037	LNSLNNPKHVLPNPSVFPNKKKEAIEAETE--NTLEN--TKILLK--HYKGLVKYNGE	1089
Qy	1261	-----QYWYFAPADALITLTKLPRVEDTVNRILNDGNTSRFIVP-----VI	1303
Db	1090	SSPLKTLSESIQTEDNYASLENPKVLSKEGKLKD--NLNLEKKKLSYLSGLGHLHIA	1146
Qy	1304	TTEQIRKNLSYFYGSGYSLSLSPYNNMIDLNL-----	1338
Db	1147	ELKEVIKNKNT-----GNSPSENNTDVNNALYESKYPFPECTDVTATVVSSEGS	1196
Qy	1339	-----VENDTWI--DVDNVNVRNITI-----RSDE-----IQKE--	1366
Db	1197	TLEQSQPKKPASTHVGAESENTITTSQNVDDVEDVDVIVPIFGESEEDYDGLQGVVTEGAV	1256
Qy	1367	--LIENILSKLNTIE-----DNKIILNNHTINFYGDINESNRFTSLTFSIL	1409
Db	1257	TPSVIDNLSKIENEYEVILYKPLAGYVRSLLKQLENNVTFVNVKD---ILNSRPNKR	1313
Qy	1410	EDINIIIEIDLVSXKYLISGNCM-----KLIENSDDIOQKID-HIGFNG	1454
Db	1314	ENPKNVLESDLI--PYKDLTSSNVVWDPYKFLNKEKRDKFLSSYNYIKDSITDINFAN	1371
Qy	1455	EHQKYIVFSYIDNETKNGPIDYSKKEGLFTAEPSPNESIIRNIYMPDSNNL-FIYSS-KD	1512
Db	1372	DVLGY--YKILSE--KYKSDLD-SIKKYINDKQGENEK-----YLPFLNNIETLYKTVD	1421
Qy	1513	LKOIRINKGDKVLLIGNYFKDDMKVLSLs-----FTIEDNTNITKINGVILDEN--GVAQI	1565
Db	1422	KIDLFIVHL-EAKVLYNTYKSNVEVKIKELNYLKTQD---KLADFKNNVFVGIADL	1476
Qy	1566	LKFMNNAKSALNTSGLNWFLESINI-----KNIFYNNLPDNTIEFILD	1608
Db	1477	-----STDYNNHNLTKFLSTGMVFENENAKTVLSNLLDGNLOGLNL	1517

RESULT 12

ID	SPOF	SCHPO	STANDARD;	PRT;	1957	AA.
AC	Q10411;	O9USE9;				
DT	01-OCT-1996	(Rel. 34,	Created)			
DT	01-OCT-1996	(Rel. 34,	Last sequence update)			
DT	15-SEP-2003	(Rel. 42,	Last annotation update)			
DE	Sporulation-specific protein 15.					
GN	SPO15 OR SPAC1F3.06C.					
OS	Schizosaccharomyces pombe (Fission Yeast).					
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;					
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;					
OC	Schizosaccharomycetes.					
OX	NCBI_TaxID=4896;					
FN	[1]					
FN	SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.					
RP	MEDLINE=20107136; PubMed=10639340;					
RP	Ikemoto S., Nakamura T., Kubo M., Shimoda C.;					
RA	"S. pombe sporulation-specific coiled-coil protein Spo15p is localized					
RT	to the spindle pole body and essential for its modification.";					
RL	J. Cell Sci. 113:545-554(2000).					
RL	[2]					
RP	SEQUENCE FROM N.A.					
RP	STRAIN=972;					
RP	MEDLINE=21848401; PubMed=11859360;					
RX	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,					
RA	Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,					
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,					
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,					
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,					
RA	Hollroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,					
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,					
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,					
RA	Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,					
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,					
RA	Seaton J., Simmonds M., Squares R., Squares S., Stevens K.,					
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,					
RA	Woodward J., Volkheart G., Aert R., Robben J., Grymonprez B.,					
RA	Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,					
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,					
RA	Borzyn K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,					
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,					
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,					
RA	Galibert F., Advies S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,					
RA	Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,					
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,					
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.U.,					
RA	Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,					
RA	Sipakovits G.V., Ussery D., Barrell B.G., Nurse P.;					
RT	"The genome sequence of Schizosaccharomyces pombe.";					
RL	Nature 415:871-880(2002).					
RL	[3]					
RP	SEQUENCE OF 705-871 FROM N.A.					

Mol. Gen. Genet. 237:359-369(1993).

[2]

RN
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
new open reading frames";
RT Yeast 9:1349-1354(1993).
CC -I- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
REPAIR.
CC
CC -I- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -I- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.sib.ch/announce/
or send an email to licensel@sb-sib.ch).

EMBL; L01992; AAA34783.1; --
DR EMML; X73541; CAA51948.1; --
DR EMML; Z28320; CAA82174.1; --
DR FIR; S38173; S38173.
DR SGD; S0001803; MLPI.
DR GO; GO:0005635; C:nuclear membrane; IDA.
DR GO; GO:0005654; C:nucleoplasm; IDA.
DR GO; GO:0006006; P:protein-nucleus import; IDA.
KW Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT FTDOMAIN 531 1678 COILED COIL (POTENTIAL).
FT FTDOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R->A (IN REF. 1).
SQ SEQUENCE 1875 AA; 21845 MW; 683AD0D34C906867 CRC64;

Query Match 3.3%; Score 290; DB 1; Length 1875;
Best Local Similarity 18.4%; Pred.No. 0.00076;
Matches 314; Conservative 304; Mismatches 612; Indels 474; Gaps 75;

QY 226 ESLSNKITANNNDIRNLKFAEDLVRLVYNQLV-----BRWNAASDIIRISMLK 278
Db :|||:::|||:::|||:::|||:
18 DERLNAIASFPGCSLEQVKSF-DGVVVGHLDKLQFNELKSENLKVTVSPDELKASSLK 76
QY 279 EDGCVVLVDVLLPCIQDFPKSTN---KPDSITTNSWEVI-----KLEAIMKYKEY 326
Db :|||:::|||:::|||:::|||:
77 KIDGLKTENVIRENDKIKRERNTFVPESVENEKOKLSSELFEFKVRKLUDDITEBKKE 136
QY 327 IPGYTSKNFDMDLEEVO-----RSFESALSSSKDKSEIFLPDL----D 365
Db :|||||:::|||:::|||:::|||:
137 TQSNOQTTLKIDERLIKEIVLRVNRRNSECKLARSTMDETAKOOGYTINDLSRTE 196
QY 366 IKVSPLVEKIAFANNVINOALSISKDSYCS-----DLVINQINKRYKIL----- 410
Db :|||:::|||:::|||:::|||:
197 LEKRKTQELTLQSNWDWLEKSLKSNEQEYLVSQRKTDKVILDIRNELNRDNDFQMERTN 256
QY 411 NDNIINFINEGTDTMTMKIFSCLKAIS-----NEDNMFMFIKITNYLKVGFAPDVSR 464
Db :|||:::|||:::|||:::|||:
257 NDVLKKQNNE-----LSKSLQBKLIIKGSLSDSLNSEKOEFSAEMS--LKQRLVDLLES 308
QY 465 TINLSGPVGVTGAYQDLLMPKDNSTIHLE-----PE---LRNFEPPTKISLTQE 514
Db :|||:::|||:::|||:::|||:
309 QLM-----AVKEELSINRETNATKVIDADSCKQPENEDLKEIQLTEKUQAQ-CER 359
QY 515 EITSLSWSFNQARAQSOFEBYKGGFYEGALGEDDNLDPAQNTVLDKDYVSVKKILLSMKT RN 574
Db :|||:::|||:::|||:::|||:
360 ECLRLSITDE-----ADEDENLISAESSSDFFILKQLIKERRI-- 399

QY 575 KEVIIH-----YIVLOQDKI-----SYEASNLFPSKDPYSILIYQKNEGSETAYYYVAD 625
Db :|||:::|||:::|||:::|||:
400 KEHLNQIEFTFIVELE-HKVPINSKFERTDMLENELNNALLHEHTSNK-----N 450

```
RL Mol. Gen. Genet. 237:359-369(1993).
[2] SEQUENCE FROM N.A.
RN MEDLINE=94205265; PubMed=8154186;
RX Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RA "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT new open reading frames.";
RT Yeast 9:1349-1354(1993).
CC -I- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC CC
CC -I- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -I- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.sib.ch/announce/ or
CC send an email to licenselib-sib.ch)
CC -----
CC EMBL; L01992; AAA34783.1; --
DR DR
EMBL; X73541; CAA51948.1; --
DR EMBL; Z28320; CAA82174.1; --
DR FIR; S38173; S38173.
DR SGD; S0001803; MLPI.
DR GO; GO:0005635; C:nuclear membrane; IDA.
DR GO; GO:0005654; C:nucleoplasm; IDA.
DR GO; GO:0006006; P:protein-nucleus import; IDA.
KW Coiled coil; DNA repair.
FT DOMAIN      69   487    COILED COIL (POTENTIAL).
FT FT          531   1678   COILED COIL (POTENTIAL).
FT DOMAIN     1834   1866   COILED COIL (POTENTIAL).
FT CONFLICT   301   301    R -> A (IN REF. 1).
SQ SEQUENCE    1875 AA; 21845 MW; 683AD0D34C906867 CRC64;

Query Match              3.3%; Score 290; DB 1; Length 1875;
Best Local Similarity 18.4%; Pred.No. 0.00076;
Matches 314; Conservative 304; Mismatches 612; Indels 474; Gaps 75;

QY 226 ESLSNKITANNNDIRNLKFADELVLRYLNQELV-----BRWNLAASDIIRISMLK 278
Db :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
18 DERLNAIASFFGCSLEQVKSF-DGDVVGHLDNKLQFNELKSENLKVTVSPFDELKASSLK 76
QY 279 EDGGVYLVDVLLPGIQPDFLKSIN---KPDSITNTSWEMI-----KLEAIMKYKEY 326
Db :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
77 KIDGLKTENVIRENDKIKRERDTFVPESVENEKMKLSSELFEFKVRKLUDDUTEKKKE 136
QY 327 IPGYTSKNFMDLDEEVQ-----RSFESALSSKDSEIFLPDL----D 365
Db :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
137 TQSNOQTLLKILDERLIKELVRENNRSNECKLARSTIMDETAKOOGYITDNLS RTE 196
QY 366 IKVSPLVEKIAFANNVINOALISLSDSYCS-----DLVINQINKRYKIL----- 410
Db :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
197 LERKTQLTLQSNNDWLEKELSKEQEYLSRKTKDKVIDLRNELNRDNDFOMERTN 256
QY 411 NDMINPSINEGTDMNTKMIFSOKLASIS-----NEDNMFMKIINTNYLKVGFPAPDVRS 464
Db :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
257 NDVLFKQNNE-----LSKSLQBLEIKGLSDSLNSKEOFSAEMS--LKQRIVLDLLES 308
QY 465 TINLSGPGVVTGAAYQDLLMPKNSTNIHLE-----PE---LRNPFPPKTI SOLTEQ 514
Db :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
309 QLM-----AVKEELNSTIREUNTAKVIADDSKKQPENEDLLKEQLTYEKUQAQ-CEK 359
QY 515 EIITSLSWFNQARAkofEBYKGFGEGALGEDDNLDPAQNTVLTDKDYVSKKILLSMKT RN 574
Db :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
360 ECLRLSSITDE-----ADEDENLSAKSSSDFI FLKKQLIKERRI-- 399
QY 575 KEVII-----YIVLOQDKI-----SVESACNFLFSKDPYSILIYQKITEGSTAT YYYVAD 625
Db :| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
400 KEHLQHQIEFTFIVELE-HKVPIINSKFETRDMLENELNNAALLHEHTSNEK-----N 450
```

Qy	626	AEIKEIDKVRIPYQI	GNKRNKIKLTF	IGHGKSEFN	TDTFANL	DV-----	DSLSS---	673					
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	451	AKVKEL-----	NAKNQKLV---	ECENDLQ	TLTKQRLC	ROIQVYLL	ITNSVS	ND	497				
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	674	-----E	IEITILNAKADIS	PKYIEIN	LLGNMPSYS	IAEBTYSG	KLKLLKORV	SEL	726				
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	498	KGPLRKEE	IQINIMQEDD	STI-----	TESDSQ	KVVTERL	VBFKNIIQL	QEKNAEL	549				
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	727	MPISQDSIT	VSANOQVEVR	INEBGRKE	ILHDHSGK	WINKBESI	IKOISSKEYIS	FNPKENK	786				
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	550	LKVVRN-----	LADKLESK-	EKSKOSQ	KIEBSETV	NEAKEALIT	LKSEKM---	DLESR	599				
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	787	IIVKSKYL	HELSTLLOE	IRNANSID	LEKKVML	TECEIN	VASNDRQI	VEGRIEEAKN	846				
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	600	IEELQELE	LELKTSP-	NEDASYS	NVTIKQ--	LTEKRD	LESQV--	QDLOTRISQ	ITR	652			
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	847	LTSDSINY	KNVEFKL	IESDLS	YDLKHQ	NGDDSH	FISFE-----	DISKTENG-	895				
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	653	ESTENS	LLNKE---	QDLYDSK	DISIS	LGKESK	SRILAEER	FKLLSN	TLNTDL	KAENDQ	709		
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	896	FRIRFINK	ETGNSIF	ETEK-	EIFSEYA	THISK-----	EISN	KDITF	DNVNGK	LVKK	947		
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	710	LRKRFYLO-	NILKQDSK	THE	TLNEYV	SKSKLSI	VE	TELLN	LKE-----	EOKLRVH	761		
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	948	VNLDAHE	VNTLSA	FFIOSLI	BYNTTK	ESLSN	LSVAMK-----	VQVYAQL	FSTG	LTNTIT	1002		
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	762	LEKNLQ	ELNKLSP-----	EKDSLR	IMVTQ	LTQKER	EDLLE	ETRKSQ	CKKIDE	LE	813		
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	1003	DASKVVEL	VS	TADETI-----	DLPTL	SEGLPI	ATI	IDG-----	DISKTENG-	895	1038		
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	814	DALSEL	KKETSQ	KDHIKOLE	EDNNSN	IEWYQNK	IEALKK	QVESVIT	SDSKOTD	IEKLO	873		
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	1039	---VSLGAA	TIKE-----	LSET-	NDPLL	RQIEAK	IGIMAV	NLTA	ASTAI-----	DISKTENG-	1078		
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	874	YKVKSE	KEFEED	KIRLHTY	NVMD	ETIND	SLAKE	LEKS-----	KINLT	DAYSOI	KEYKDL	929	
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	1079	--VTSAL	GIASGFS	ILLVPL	AGIS	PIPSLV	NNELIL	QDKAT	VIYFKHI	SLAETEGAP	1136		
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	930	YETTSQ-----	-----	SLQQT	NSKLDES	FKDF-----	FNQIK	956					
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	1137	TLDDKIL	IMPQDD	LV	SEIDFN-	NNSITL	GKCEI	BRWAEGG	SGHTL	DDIDHP	FSSPSITV	1195	
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	957	NLTDEK	TSLEDK	ISLLEK	QFN	LN	NELDLQK-----	987					
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	1196	RKPWL	SIYD	VLNTIK	KEKIDF	SKDML	PNAPNR	VFGEM	GTGPF	SR	LDN-----	1245	
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	988	-----	KGMEK	KADFK	KRISIL	QNNKE	VA	KFSEY	ESKLSKI	QNDL	DQQT	IYAN	1037
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Qy	1246	-----	DGTKL	DR	TRDH---	YEGOFY	WRYP	AFI	ADALIT	LK	KPRYED-----	1284	
		: : : :	: : : :	: : : :	: : : :	: : : :	: : : :						
Db	1038	TAQNNYE	QELQ	HADV	SKTISER	REQHTY	KGVG--	KTLN	SRDQ	LENAL	KENEKES	WSSQ	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:31:53 ; Search time 40.625 Seconds
(without alignments)
7187.047 Million cell updates/sec

Title: US-09-126-816B-6_COPY_1_1700
Perfect score: 8677
Sequence: 1 MNLVNAQLQKVMVVKFRQ.....YLYGIDRYVNVKIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PTC_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_PUBCOMB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PTCUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	6733.5	77.6	2366	15	US-10-011-366-10
2	4458	51.4	2710	15	US-10-011-366-6
3	501.5	5.8	3169	15	US-10-114-170-257
4	310	3.6	2184	12	US-10-304-095-6
5	309	3.6	1639	15	US-10-087-464-10
6	288.5	3.3	1948	12	US-10-032-585-7611
7	286.5	3.3	1610	15	US-10-155-533-9
8	284	3.3	1881	12	US-10-032-585-7646
9	268.5	3.1	2167	10	US-09-801-368-56
10	266	3.1	1847	12	US-10-032-585-7538
11	257	3.0	6281	9	US-09-815-242-12996
12	256	3.0	2285	10	US-09-932-183A-2
13	253	2.9	1295	10	US-09-726-949A-1
14	250	2.9	4563	11	US-09-802-640-32
15	249	2.9	1706	9	US-09-864-761-46862

16	249	2.9	4563	10	US-09-870-759-128	Sequence 128, Appl
17	249	2.9	4563	12	US-09-751-708A-128	Sequence 128, Appl
18	247.5	2.9	3158	9	US-09-815-242-12611	Sequence 12611, A
19	247	2.8	1805	11	US-09-820-843A-73	Sequence 73, Appl
20	245.5	2.8	2437	9	US-09-815-242-5834	Sequence 5834, Ap
21	241	2.8	1786	10	US-09-742-096-3	Sequence 3, Appli
22	240	2.8	2871	15	US-10-146-473-41	Sequence 41, Appl
23	239.5	2.8	1169	12	US-10-241-596-20	Sequence 20, Appl
24	239.5	2.8	2025	9	US-09-815-242-5703	Sequence 5703, Ap
25	236.5	2.7	1979	15	US-10-205-823-419	Sequence 419, Appl
26	233.5	2.7	3878	12	US-10-080-608A-11	Sequence 11, Appl
27	233.5	2.7	3899	15	US-10-171-311-4	Sequence 4, Appli
28	233.5	2.7	3907	15	US-10-171-311-2	Sequence 2, Appli
29	233.5	2.7	3917	15	US-10-171-311-8	Sequence 8, Appli
30	233.5	2.7	3925	15	US-10-171-311-6	Sequence 6, Appli
31	232.5	2.7	2119	12	US-09-769-744A-28	Sequence 28, Appl
32	232	2.7	1598	12	US-10-032-585-7766	Sequence 7766, Ap
33	232	2.7	1781	10	US-09-995-749A-2	Sequence 2, Appli
34	229	2.6	1107	12	US-10-032-585-7483	Sequence 7483, Ap
35	228	2.6	2659	12	US-10-311-879-28	Sequence 28, Appl
36	227	2.6	1240	12	US-10-032-585-7366	Sequence 7366, Ap
37	227	2.6	1288	15	US-10-006-780-2	Sequence 2, Appli
38	224	2.6	1270	12	US-10-032-585-7127	Sequence 7127, Ap
39	220	2.5	3788	11	US-09-952-267-76	Sequence 76, Appl
40	219	2.5	1818	12	US-10-032-585-7616	Sequence 7616, Ap
41	218	2.5	3092	10	US-09-801-368-172	Sequence 172, Appl
42	217.5	2.5	1978	12	US-10-094-466-64	Sequence 64, Appl
43	217.5	2.5	2893	12	US-09-882-227-522	Sequence 522, Appl
44	217	2.5	2789	9	US-09-801-574-57	Sequence 57, Appl
45	216.5	2.5	1482	9	US-09-815-242-12484	Sequence 12484, A

ALIGNMENTS

RESULT 1
US-10-011-366-10
; Sequence 10, Application US/10011366
; Publication No. US20030054493A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/10/011,366
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-NO. US20030054493A1-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:
 NAME: Ingolia, Diane E.
 REGISTRATION NUMBER: 40, 027
 REFERENCE/DOCKET NUMBER: OPD-01121
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2366 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Query Match 77.6%; Score 6733.5; DB 15; Length 2366;
 Best Local Similarity 76.4%; Pred. No. 0;
 Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;

QY 1 MNLVKAOLQKVVYKFRIOEDYVAILNALAEYHNHNMSSVVEKYLKLDINLTDNYL 60
 DB 1 MSLVNRKQLEKMANVFRQEDYVAILDALAEYHNHNMSSVVEKYLKLDINLTDNYL 60

QY 61 NTYKSGRNKALKKFKYELTMEVLEKLNLSLTPVEKNLHFIWIGQINDTAINYNQWKD 120
 DB 61 DTYKSGRNKALKKFKYELTMEVLEKLNLSLTPVEKNLHFIWIGQINDTAINYNQWKD 120

QY 121 VNSDYTVKPYVDSNAFLINTLTKTIVTESATNTNLESFRENLDNDPEPYNKFVKRMELIY 180
 DB 121 VNSDYTVKPYVDSNAFLINTLTKTIVTESATNTNLESFRENLDNDPEPYNKFVKRMELIY 180

QY 181 DKQKHFDYKSOIENPEFIDNIITKYLSEYKDEALNKYIIESLNKLTANNNDI 240
 DB 181 DKQKNPINTYKQREENPELIDDIIVKYLSEYKDEALNKYIIESLNKLTANNNDI 240

QY 241 RNLEKFAEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFS 300
 DB 241 RNLEKFAEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFS 300

QY 301 INKPSITNTSMEMIKLEAIMKYEIPGYTSKNFDMLEDEEVQSPESALSKSKSEIF 360
 DB 301 IEKPSVTVDFWEMTKLEAIMKYEIPGYTSKNFDMLEDEEVQSPESALSKSKSEIF 360

QY 361 LPDLIDKVPLEKIAFANNSVINOALISLKSQSDIIVNOIKNRYKILNINLPSINE 420
 DB 361 SLSGDEASPLEKIAFANNSVINOALISLKSQSDIIVNOIKNRYKILNINLPSINE 420

QY 421 GTDFNTMTKIFSDKLASIGNEDNMFMWIKITNLYKVGFAFADVRSTINLSGPGVYTGAYQD 480
 DB 421 DNDFTNTTTFIDSINVAENADNGRFWELGKYLVRGFPDVKTTINLSGPEAYAAQD 480

QY 481 LMFKNQSNHILPELARNPPKTKISQLEQETISLWSNQARAKSQFEYKKGYPE 540
 DB 481 LMFKEGSMNHLIEADLENFEISKTNIQSQTEQENASLWSPDARAKAQFEYKKNYPE 540

QY 541 GAGEDNDLDFACQTVLKDQYKSKTILSSMKTENKEYIHYIYVLOQDKISYEASCNLFSK 600
 DB 541 GSIGEDNDLDFSQNIIVVDKEYLEKISLSARSEYIHYIYVLOQDKISYEASCNLFSK 600

QY 601 DPYSILYQNKIBEGSETAYVYVADAEIKEDIRIPYQISNKNRIKLTFIGHKGSEFNT 660
 DB 601 TPYDVLFOKNIEDSIAYYYPGDEIGQIDKYKIPSIISDRPKILTFIGHKDEFT 660

QY 661 DTFANLDVDSLSSEITILNLAADISPKYIEINLGCNMFYSYIYAETYPGKLLKTK 720
 DB 661 DTFAGDVSLSSEIEAIDLAKEDISPKYIEINLGCNMFYSYIYAETYPGKLLKTK 720

QY 721 DRVSELMPSISQDSITVSANQYEVNRNEKREILDHSGWINKESIIKDISKEYISF 780
 DB 721 DKISELMPSISQDSITVSANQYEVNRNEKREILDHSGWINKESIIKDISKEYISF 780

QY 781 NPKNKIIIVKSKYLHELSTLLQEIIRNANSDDILEKKVMLTECEINVASNIDRQIVEGR 840

DB 781 NPKNKIIIVKSKYLHELSTLLQEIIRNANSDDILEKKVMLTECEINVASNIDRQIVEGR 840

QY 841 IEAKNLTSDSINYIKNEFKLIESDSLYDLKHQGLDDSHFIFEDISKTEGFRIRF 900

DB 841 IEAKNLTSDSINYIKNEFKLIESDSLYDLKHQGLDDSHFIFEDISKTEGFRIRF 900

QY 901 INKETGNSIFETEKEIPEYATHISKEISNIDTIFDNVNGKLVKKNLDAHVEVNTLN 960

DB 901 INKETGESIFVETEKTIPEYANHITTEISKIKGTFIDTVNGKLVKKNLDTTHEVNTLN 960

QY 961 SAFFIQSLIEYNTTKESLNSLVAMKVQVYAOFLSTGLNTITDASKVELVSTALDEITD 1020

DB 961 AAFQIQLIEYNSKESLNSLVAMKVQVYAOFLSTGLNTITDAAKVELVSTALDEITD 1020

QY 1021 LLPTLSEGLPIATIIDGVSLGAAIKELSETSDPLLRQIEAKIGIMAVNLTAATTAIT 1080

DB 1021 LLPTLSEGLPIATIIDGVSLGAAIKELSETSDPLLRQIEAKIGIMAVNLTAATTAIT 1080

QY 1081 SALGTASGFSILLVPLAGISAGIPSLVNNELTLODKATKVIDYFKHISLAETEGAFLLD 1140

DB 1081 SLSGTASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVIDYFKHISLVETEGVFTLLD 1140

QY 1141 DKIMPODDLVLSEIDFNNSITLGCBIWRAEGSGGHTLTDDIDHFFSSPSITYRKPL 1200

DB 1141 DKIMPODDLVLSEIDFNNSITLGCBIWRAEGSGGHTLTDDIDHFFSAPSITYREPHL 1200

QY 1201 SIYDVNTKKEIDPSKOLMVLNPNRVFGVEMGWTGFRSLDNDGTYKLLDRIDHVEG 1260

DB 1201 SIYDVLEQBELDLKOLMVLNPNRVFAWETGTPGLRSLDNDGTYKLLDRIDHVEG 1260

QY 1261 OFYWEYFAFIADALITLKPVEDTNVRLNDGNTSRFIVPVTTEQIRKNLSYFYSGG 1320

DB 1261 EYWFYFAFIADALITLKPVEDTNVRLNDGNTSRFIVPVTTEQIRKNLSYFYSGG 1320

QY 1321 GSYSLSPYNNIDNLVENDTWVIDVNVVKNITTESDEIQKGLIENILSKLNIEDN 1380

DB 1321 GTYALSLSQYNGINIELSESDVMIIDVNVVDRVTIESDKIKGDLIEGILSTLSIBEN 1380

QY 1381 KIILNNHTINFGVDNENRFLISLTFISLEIDNIIIEIDLVSYSKILLSGCMKLIENS 1440

DB 1381 KIILNSHINFSGEYNGSVFSLTFISLEGINAIEVDLLSKYSKILLSGELKILMLNS 1440

QY 1441 SDIOQIDHIGFNGHQKVIYFVSDNETKNGFTDYKSGELFTAEFSNESIIRNIYMP 1500

DB 1441 NHIQKIDYIGFNSLQKNIPIYSFVDSKENGFTNGSTKGLFVSELPDVVVLISKVMD 1500

QY 1501 DSNNLFIYSSKDLDIRINKGDVKKLLICNYFKDDMKVLSLFTIEDTNTIKLNGVYLDEN 1560

DB 1501 DSKPSFGYNNLKDVKITKDNVNILGYLKDIIKISLSLTLQDEKTIKLSVHLDES 1560

QY 1561 GVAQILKFMNNAKSALNTSNLNFLESINIKNIFPNNDPNIPILOTNFIISGNSIG 1620

DB 1561 GVAELKFMNR-KGNTNTSDSLMSFLESNMKISFVNFQSNIKFILDFANFIISGTTISG 1619

QY 1621 OFELICDCKDKNIQPIYFNFKIETSYTYLVGNRQNLIVPEPSVHLDDSGNISSTVINFSOK 1680

DB 1620 QFEFTCDENDNIQPIYFNFKIETSYTYLVGNRQNLIVPEPSVHLDDSGNISSTVINFSOK 1679

QY 1681 VLYGIDRVNVKVIAPNLYT 1700

DB 1680 VLYGIDSCVKNVISPNIYT 1699

RESULT 2

US-10-011-366-6
 ; Sequence 6, Application US/10011366
 ; Publication No. US20030054493A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Williams, James A.
 ; Kink, John A.
 ; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
 ; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE

DISEASE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESS: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/011,366
FILING DATE: 16-Dec-1992
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/957,310
FILING DATE: 23-Oct-1997
APPLICATION NUMBER: US 08/329,154
FILING DATE: 24-Oct-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-Dec-1993
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-Dec-1992
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-Oct-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingollia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Query Match 51.4%; Score 4458; DB 15; Length 2710;
Best Local Similarity 50.7%; Pred. No. 2.2e-270;
Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;
QY 1 MNLVNAQLQKVVYKFRIOEDYVAILNALBEEYHNMSESSVVEKYLKLDINNLTDNYL 60
DB 1 MSLISKEELIKLAY-SIRPRENEYKTLTLDYENKLTNNENKYLQKLKLNESIDVPM 59
QY 61 NTYKSGRNKALKKFKKEYITMEVLELKNNSLTPVPERKHLFIWIGQINDTAINYNQKD 120
DB 60 NKYKTSRRNALSNLKKDKILKEVILIKNSLTPVPERKHLFIWIGVSDIALEYIKQWAD 119
QY 121 VNSDYTVKVFYDSNAFLNLTWLTCTIVESATNTLESFRENLDNPFEDYKFKRMEIY 180
DB 120 INAEYNIKLYWDSAEFLVNLTKKAIYESSSTTEALQLLEBEIQNPQDNMKFYKCRMEY 179
QY 181 DKQKHFDYKQIEENPEFIDNIITKYLNSYKDLKALNKYIEESLNKITANNNDI 240
DB 180 DKQKRFINTYKQINQKPTVPTIDDIILKHLVSEYNRDETVELSYRNSLRKNSNHGIDI 239
QY 241 RNLEKFAEDLVRLYNQELVERWNLAASDIARISMLKEDGGVYLDVILPGIQDLPFKS 300
DB 240 RANSLFTEQELLNYSQELLNRGNLAASDIARVLLALKNFQGVYLDVDMLPGLHSDLPKT 299
QY 301 INKPSITNTWEMIKLEAIMKYKEVPGYTSKNFMDLBEVORSFESALSKDSKEIP 360
DB 300 ISRPSSIGLDRWEMIKLEAIMKYKYINNYTSENFDKLDQQLKDNFKLAIIESKSEKSEIF 359

QY 361 LPLDDIKVSPLEVKIAPANNVINQALISLSDSYCSDLVINQIKRYKILNDNLNPSINE 420
DB 360 SKLENLNSDLKIAFALGVSINQALISKOGSYLTNLVIEQVKNRYQFNLQHLNPAIES 419
QY 421 GTDFNTWKIFSDKLASISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTCAYOD 480
DB 420 DNNFTDTTKIFHDSLFNSATAENSMTLKIAPIYQVGFMPPEARSTISLSGPGAYASAYD 479
QY 481 LLMFKDNSTNHLLEPRLNFFPKTKISQLETEITSLWSFNQARAKSQFBEYKKGYPE 540
DB 480 FINLOENTIEKTLKASDLIEFKFEPENNLSQLEQINSLSWPDQASAKYQFKEYVRYDTG 539
QY 541 GALGEDDNLDPQNTVLKQY-VSKILSS--MKTRNKYIHYIYVLOQGDKISYEASCNL 597
DB 540 GSLSDNGVDNFKNTALDKYLLNNKIPSNNVEEAGSKNYVHYIYVLOQDDISYEATCNL 599
QY 598 FSKDPYSSTLYQKNEGSETAYYYVAD--ASIKEDKYRIYQISNKNENIKLFTIGHGK 655
DB 600 FSKNPKNSIIQRMN--NESAKSYFLSDGDESILELNKYRIPERLKNKKEVKVYTFIGHGK 657
QY 656 SEFNTDTPANLDVDSLSSEIETILNLAADIAPKYEINLLGNMFSYSIYAEETYPGKL 715
DB 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLDISPKNVEVNNLLGNMFSYDFNVEETYPGKL 717
QY 716 LLKIKDRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWINKKEEIIKDISK 775
DB 718 LLSIMDKITSTLPDVNKNISITIGANQYEVRIINSEGRKELLAHSGKWINKKEEIMSDLSK 777
QY 776 EYISFNPKENKIIIVSKYLHELSTLLOEIRNNANSDDILEKKVMLTECEINVASNIDRQ 835
DB 778 EYIFFSDSNKLUKAKSKNIPGLASISEDIKTLLDASVSPDTKFTLNNLKLAIESSIGDY 837
QY 836 IVEGRIEERAKNLTSDSINVYKNEFKLIESISDLYDLKHQNGDLDSDHFSFEDISKTENG 895
DB 838 IYKLEPVRKNIIRNSIDDLIDEPNLENVSDLELKLKLNLDKELISFEDISKNST 897
QY 896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHE 955
DB 898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLLDNIQLDHTSQ 957
QY 956 VNTLNSAPFIQSLIENYNTKESLSNLSVAMKVQVTAQLFSTGLNTITDASKVVELYSTAL 1015
DB 958 VNTLNAAPFIQSLIDYSSNKKVLDLSTSVKQVLAQLFSTGLNTIYDLSIQLVNLISNAV 1017
QY 1016 DEFIDLLPTLSGLPIIATIIDGVSIAGAIKELSETNDPLLRQETEAKEIGIMAVNLTAAS 1075
DB 1018 NDTIINVLPTITGPIVSTILDGINLGAALKELDEHDEHLLKKELEAKGVLAINNLSLI 1077
QY 1076 TAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGA 1135
DB 1078 AATVASIVGIGAEVITFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNHLSKKYGP 1137
QY 1136 FTLLDDKIIMPODDLVLSEIDPNNSITLGKEIWRABEGSGHTLTDDTDHFFSPSPITY 1195
DB 1138 LKTEDDKILVPIDDLVISEIDFNNSIKLGTGNILAMEGGSGHTVTGNDHFFSPSPISS 1197
QY 1196 RKPWLSIYDLNKKKEKIDFSDKLVLPNAPRVFCYENGWTPGERSLNDGCTKLLDRIR 1255
DB 1198 HIPSISYSAIGIETENLDFSKIMWLPNAPRPFVWETGAVPGLRSLNDGCTKLLDRIR 1257
QY 1256 DHYEQVYRYFAPTADALITKLPRYEDTVNRINLDGNTRSFIVPVIITEQIRKNLSYS 1315
DB 1258 DLYPKFYRWFYAFP-DYAITTLKPYVEDTNKIKLDDKTRAFIMPTITITNEIRNKLVS 1316
QY 1316 FYSGGYSLSLSPYNNIDNLVENDTWVIDVDNVVKNITTESDEIQGELIENLSKL 1375
DB 1317 FDGAGGTSLLSYSPISNTINLSKDDLWIFNIDNEVEISIENTGKIGKLIKDVLSKI 1376
QY 1376 NTEEDNKILNNTINFYGDINESNFISLTFSILEDINIIIEIDLVSKEYILLSGNCWK 1435
DB 1377 DINKKLIIGNOTIDFSGDIDNKDRIYIFLTCLDDKISLIIEINLVAKSYLLSGDKNY 1436
QY 1436 LIENSDDIQKIDHIDGFNGEHQKIFYSYID-NETKYNFIDYSKKEGLFTAESNESII 1494


```

Db 1437 LISLSNTTEIKNTGLD---SKNTAYNTDESNNKYFCAI-----SKTSQKSII 1483
QY 1495 RNIYMPDSNNL-----FIYSSKDL--KDIRINKGVKLLIGNYFKD---DMKVSLS 1541
Db 1484 H--YKDSKNILEFYNDSTLEFNSKDFIAEDINVFMKODINTITGKYVDNNTDKSIDFS 1541
QY 1542 FRIEDNTIKLGVLDENGVAQLKFMNNAKSALNTSNLNFLESINIKNIFYNLDP 1601
Db 1542 ISLVSKQVKNGLYNLVSYSYDFVKNSDGHHNTSNFMLFLDNISFWKLGFGE--- 1598
QY 1602 NIEFLDNTFIAGSNGSIQFELICDKDKNIOPYFINFKIKETSITYLVGNRQNLIVEPS 1661
Db 1599 NINFVIDKYFTLVGKTNLGVFEICDNNKNIDYFGEWKTSKSTIFSGNGRNVVERPI 1658
QY 1662 YHLDGSGNSTVINPFSQKLYGIDRYVKNVJIAPNLYT 1700
Db 1659 YN-PDTGEDISTSLDPSYBPLYGIDRYINKVLIAPDLYT 1696

```

RESULT 3

```

US-10-114-170-257
; Sequence 257, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod

```

TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/114,170

FILING DATE: 01-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/453,702

FILING DATE: 03-DEC-1999

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 257:

SEQUENCE CHARACTERISTICS:

LENGTH: 3169 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 257:

US-10-114-170-257

Query Match 5.8%; Score 501.5; DB 15; Length 3169;
Best Local Similarity 20.1%; Pred. No. 4.6e-22;
Matches 421; Conservative 331; Mismatches 723; Indels 619; Gaps 104;

```

QY 14 VVKPRI---QDEYVAIINAL--BEYHNMSSESSVVEKYLKLDKINLNTDNYLNTYKSGR 68
Db 160 YIKIRKTRGAEDQTTTQSLLIINELNGVDNRNTI-PEQKISELNDIHSYENMQIKNSR 218
QY 69 N--KALKFKFEYLWELVLELKNSS-----LTPVEKNLHFTWIGQ 106
Db 219 KGIETLVKQGLSSLINDNKGKQSLDASKIINLLGIEYQSHKVDIEPIHAWVAGA 278
QY 107 INDPAINTVINOQVNSDYTVKVFVDSNAF-----LINTUKKT----- 144
Db 279 PPDNTFSYITAFNTYKDYTYLLWDPNAPGAAKFSGILKNIAMYAIIMRLRRTNPHLAE 338
QY 145 -----IVESATNNTLE--SPRENLDNPEFY-----NKFYKMEIILYDKQKFI 187
Db 339 EMNEVILKIQIONETIEFKETRERKLELENRYKSLTSEKFKFNVFLESIMGMQDNYF 398
QY 188 DYKYSQIBENPEFI-----IDNIIK--TYLSNEYSKDLEALNKYIESLNKITANGN-- 238
Db 399 TYCISNGISNTDDISRDLFTNLVLSPEVQDNDFKSTVEKNKRDIDLKNTISQKFGDRP 458
QY 239 ---DIRNLEKFADEDLVRLNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQP 295
Db 459 QLRDINTLESFKKPDYFYQOEMLLRWNYAAASQVRAINILKEYGGIYTDITLIPAYSD 518
QY 296 DLFPSINKPDSITNTSWEMIKLEAIMKY--EYIPG--YTSKNFPMLEEVQSRSEFALS 351
Db 519 KVSQIINE-KSDDKRFEDLKRRIISIESILSLKGEKYSIKH-DGLDETTLNQNLNL- 575
QY 352 SKSDKSEIFLDDIKVSPLEVKIAFANNSVI-----NOALISLKDYSVC 395
Db 576 SEIEK-----LIDDY-FKPVEKVVDRDFKFKYQKWTENTWIRGNPNMLTHGSKC 630
QY 396 SDLVINQIKRYKILNDNLNPSINEGTFNTMKIFSDKLASISNE-----DNM 444
Db 631 IDFILSGQKQYLEL-QRIRDNISYNNLFYTT-----EDLKSLLNVAIGGIPAKKYLEHG 684
QY 445 MEMIKITNLYKVGFPADYRSTINLSGP----- 471
Db 685 LF-----SEYRQDGTIPYVVSTLNSISGPDIMRQMKYKYSIGRIGEVHIKONKSDVNPL 740
QY 472 GYVTGAYQDLLMF---KDNSTNIHLEPE-----LRNFEPPKTKISQLT---BQEITS 518
Db 741 GYVASSNKDNKSFNWLNPVSVGINDITDDDESSWAVRNNDINKILFEKINCHVPEKLPTS 800
QY 519 LWSFNQARAKSQFEEYKGYFEGALGEDDNDLDFQAQNTVLDKDYVSKKILSSMKTRNKEYI 578
Db 801 LY-----YETDSRSFFQGW---DNKSIKHVTETINKDLI--KQINLLLTSSNIDV 844
QY 579 HYIVLOQGDKISYEASCNLFKDPYSSITLYOKNEGSETAYYYVYVADAEIKEDKYRIPY 638
Db 845 KLLIKL--DRELYAISKI-----DNPLAURISRTLOQLANVYVTSNTFEPENTINFIY 896
QY 639 QISNKRN-----IKLTFIGHGKSEFNFTDTFANLDVDS-----LSSEIET 677
Db 897 DFKKQDDLLSAIKL-----FSRNDADTKIIVWYNSVMKXVFLREVISC 942
QY 678 IINLAKADISPKYIEINLLGCNMFYSYIAYETPGKLLKIKORVSELMPSISQDSI-- 735
Db 943 VLRSKKVD---SYINEN-----KKNLSKEDA--GALRDYAKLKWELFSLMLDDGYYK 990
QY 736 TVSANQYEVRIINEGKREILDHSGKWINKESIIDKISSKEYISFNP-----K 783
Db 991 IITTNAY---IKERDKL-----SGIYNIENSIISGHESFDIIRSNQHWGLDSTVEQFK 1042
QY 784 ENKIIVKSKYLHELSTLLQEIIRNNANSDDIDLEKKVMLTECEINVAS-----NIDRQIVE 838
Db 1043 KPEFYVKSE-LSSAKSIFFDIKNK-YITDPETKRNVLVHQLDSDDIKERIAFLDISHYAP 1100
QY 839 GRIEAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHQNGL 878
Db 1101 GSLLEKQLQS---GYVPSDINIIEAYLLASVSGHYSHGVVYPAPSKLLELRRTK 1156

```


Db 1141 REKKDKNIKKRYKKNKFNPNVSLNNICNFSLKCLGNMRHNNLSLFKNLTTLTKTGEIELKL 1200
QY 575 KEVIHYIV-----QLOGDKISYEASC-----NLFS--KDPY 603
Db 1201 KWLHLKWNFYKKGKMKKYIRNKLKNNKKIYAYICIGDFSNCYEHNNYLFKILKNFF 1260
QY 604 SSI-----LYOKNIEGSETAY-----YVADAEIKEIDKRYIPYOI 640
Db 1261 DNINNEFEIYLPKRSFRLYNKLNNSFLSYYPNVKSGFLHY-----IRNLRELIKSHL 1315
QY 641 SNKRNIKLTFIHKSEFN-----TDTFANLDVD--SLSSEIETILNAKADISPKYIEIN 694
Db 1316 NDNHFLQMPKTKSKSLYIFADSYKSLQVDRDIFMTIITVIRYVYLVNYSFISKEPK 1375
QY 695 LLGCNMFYSYIAEETYPGKLL-LKIKDRV-----SELMPSISQDSITV 737
Db 1376 LARKNIFYFOIENOMKGVYLSVRDKRVENIKWYLSNMKKINHDEILESKNSSINI 1435
QY 738 SANQVEVRINECKREILDHSGKWINKESIITK-----DISKEYISFNPKENKII 788
Db 1436 NKNFMICINHEQDTE--EKGNTQNKERKHDYIGPIYNNSPDSTTTTHSSNNYKGNH 1492
QY 789 VKSKY-----LHELSTLQE-----IRNANSDDIDLEKVMLTECHINVASNIDRQI 836
Db 1493 VSGDYKNDGLLHKGNNSMNECYVKDKCNMNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1552
QY 837 VEGRIEAKNLSDSI--NYIKNEFK--LISISDSLYDLKHQGLDP---SHFISPED 888
Db 1553 I---IKYHKTIDTDSKNHNTYFNKFLNFDLKKIISNIYGLPQGSLSNLCSLYAYLD 1609
QY 889 ISKTENGPRIRFINKETGNSIETEKETFSVATHISKEISNIKDTIDFNVNGK---- 943
Db 1610 --KNEEFQNLVSEKQINKNYFLANGT---CHYFNLSLILFIDDFLITLKNKNKIF 1664
QY 944 ----LVKKYNLDAAEVNTLSNAPFIQSII-----EYNTTKESLSNLSVAMKV 987
Db 1665 KNLLLKKKI---WGSNINSKTKFKIPLIYKNDLLIYNFQNKYQKKYKIRKXKQISV 1721
QY 988 ---QVYAOQFSTGLNTITDASKVELVSTALDETDL--PTLSEGLPIIA--TIIDGVS 1041
Db 1722 RNRKIRHQLVNAKKKHTSVQK-----DKINKYINLIHPTIQKNDVLSNNSINWFERI 1775
QY 1042 GAAIKELSTNDPFLRQETEAIGIMAVNLTAASTAIVTSALGIASGPSILLVPLAGISA 1101
Db 1776 --YIKESHKSN-----SSIRT 1789
QY 1102 GIP-SLVNNELIQDKATKVIDYFKHISLAETEGAFLLDDDKIIMPQDVLVSEIDFNN 1160
Db 1790 DIPNSVNDDEIYVQKSD-----NNSYSTNNLYNNINMTQNG-----DNN 1829
QY 1161 SITLKGCEIWRAGGSGCHLTDDIDHFFSPSITVRKPMWSIYDVNLINKEIKIDFSKOLM 1220
Db 1830 NVNIEK-----HVQNDSDFOCFNSN-----NLYIEKDIKENNI----- 1861
QY 1221 VLPNAPNRFVGYEMGWTPGFRSLDNDGTLKLDRIRDHYBGQFYWRFYAFIADALITKLKP 1280
Db 1862 ---SQINRKLCKRNFKSKRKINTLYLQIKV-----IKLKCKK 1900
QY 1281 RYEDTNVRINLGNTRSFIVPVITTEQIRKNLSYGFYGGSGYSLSLSPYNNMI-DNLV 1339
Db 1901 KYIKHKKMKYMNQNF-----KKLKKLQKF---HNASFELKINKINKIRLNKL 1949
QY 1340 ENDTWIDVDNVVNKNTIBESBIOKGELIENILSKLNIEDNKIILNNHTINIFYGIDNESN 1399
Db 1950 KKR-----KNHSI-----NITPVTSIB---WLANSYTFDP---INNS- 1980
QY 1400 REISLTFSLIEDNIIIEIDLVSYSKILLSGNCKMIENSDDIQKIDHIGFNGEHOKY 1459
Db 1981 -----IQSTYP--WKMKCATIRNHLHNVI-----IDKKNKY 2014
QY 1460 IFYSYIDNETKNGFIDYSKSGELFTAFESNBSIIRNIYMPDSNNLFYSSKDLDIRII 1519
Db 2015 FMKNLVENRIVRNIISKQKCSQLYKQK-----QNVYFCYKNNFSLKSSILKFIICI 2067

QY 1520 NKGDVKLLIGNYFKDDMKVSLSFSTIEDTNTIKLNGVYLDENGVAQILKFMNNAKSALNTS 1579
Db 2068 -----KTLKKOFNAFT-----NSTY-----NTRKIL--- 2088
QY 1580 NSLNFLESINIKNIFNNLDPNIEFILDITNFIISGNSISGOFELICDKDKNIQYFNF 1639
Db 2089 -FLISYMNKMLIKN-----KLLKFV--KLFLQTAIEAFYARIFNQQDSFYPCLOHF 2138
QY 1640 -KIKETSYTLY-VGNRQNLIVEPSYHLDDSGNISSTVINFQSKYLYGDRYVNVK 1692
Db 2139 RKIKKRLINKYKIGHKNKLLREFFP-----LENFIKKELYNWPYMFKI 2182
RESULT 5
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chieshti, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; PRIORITY FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-10

Query Match 3.6%; Score 309; DB 15; Length 1639;
Best Local Similarity 19.5%; Pred. No. 2e-10;
Matches 351; Conservative 280; Mismatches 554; Indels 618; Gaps 94;
QY 36 NMSESSVVEKYKLLKLDINNLDNTYNTYKSGRNKALKKFKEYLTMEVLELKNLSLTPVE 95
Db 112 NPSSSSSDSAKSVADLKHVRVNYLLT-----IKELYPQDFDLTNHMLTLC 159
QY 96 KNLHFITWGGQINDTAINY-INQMKVNSDYTVKPVYDSNAFLINTLTKKTIVESATNNTL 154
Db 160 -NIH-----GPKYLDIGYEEIN-----ELLYKLN-FYFDLRAKLVNDVCANDYC 201
QY 155 E-SPRENLDPEFDYK-----FYKRMELIYDKQKHFDYDYKSGQIEENPEFIIDNIKTY 209
Db 202 QIPENLKITRANELDVLKLVFGYRKLPLDNIKDNVGMEDYIKKN----- 245
QY 210 LSNEYSKDLKALNKYIEBSLANKITANNNDNRNLEK-----FADEDLVRLYNQELVERN 265
Db 246 -----KKTIEINELIEES-KKTIDKKNATKEEKKLYQAQYDL-SYNNKQEEAHNL 298
QY 266 AAASDILRISMLKEDGGVYLDVILPGIQ-PDLFKSINKPDSITNTSWEMIKLEAIMKYK 324
Db 299 LSVLE-KRIDTLKKNENIKELDKINEIKNPPANSNGTNPNTL-----LDKKNKIE 348
QY 325 EYIPGYSKTFMDLDEEVQORSFESALSSKSKSEIFLPLDDIKVSPLEVK--IAPANNVS 382
Db 349 EH-----EKEI-----KEIAKTIKFNIDSFLTDPLEYLEYREKKNKI 386
QY 383 INQALISLSDSVCSDLVNIQIKN--RYKILNDNLNPSINEGTDFNTMTKIFSDKLASISN 440
Db 387 DISAKVETKES-----TEPEYFNGVTYPLSYNDINNALNLSFG----- 427
QY 441 EDNMFMMIKITNYLKVGFAPDVRSTINISGPGVYTGAYQDLLMFKDKNSTNIHLEPELRN 500
Db 428 -----DLINPFDYT-----KEPSKNIY-TDNERKK 451

QY 328 PGYTSKNFMDLDEEVORSFESALSSKSKSEIFLPLDDIKVSPLEVKIAFANNSVINQAL 387
DB 667 KTLMDQ-LESTGPHFVRCILPNLEKRAKFDKNVLGQDRC-----NGVLEGR 714
QY 388 ISLKSDYSCDLVINQIKRYKILNDNLNPSINEGDTFNTMTKIFSDKLASISNEDNMFM 447
DB 715 IT-RAGYPNRMFDEFIQRYSIICDELSSPQKNTNCEIL-----KFVKLAPDPFKVL 768
QY 448 IKITNVLKVGFPADPVRSTNLGSGPVYTCAYODLLMFKDNSTNIIHLELPELRFEPKTK 507
DB 769 TKI--FFKNGILKGLKLEIIRDLAKNIFTDLQK---VIRGNLRLVL-----KQK 812
QY 508 ISOLTEQETLSLWSFNQARAKSQFEBYKGYFEGALGD-----DNLDPAQN---TVLD 558
DB 813 IKIQAQIILSRWTWLTDEIKNSPMRLFFHVHPLEDSAKVLDSKQLOENLQTLTVKL 872
QY 559 KDYVSKILSSMKTRKEYIHIVIQLGDKISYEASCNLFSDPYSSILYQKNIQIEGSETA 618
DB 873 KD--SEKLTGLETNDEKLEKREQMNNLODEMINTSIK--EKD---DKLSQLRLTENKSK 925
QY 619 YYYVADAEIKEDKY-----RIPQISNKRNIKLTFTGHGKSEFWT-----DT 662
DB 926 HRIETLEIKLDFEQKQNDLINEHEKLTTRSLDELHDKHNAK-----EELNSLKLHD 978
QY 663 FANLDVDSLSSSETILNLAKA-----DISPKYIEINLGCNM-----700
DB 979 SAQSELSNKKQKLEITKLNQAHSKELLKFKQMHDKSIEDLNEKVSRLNKKLQADLN 1038
QY 701 FYSYIYAETYPKGL---LLKIKDRVSELMPSISODSITVSANQVEVRINEREGREILDH 757
DB 1039 SSKDVISSE--HGKLQNEIMKLKELCSDYERKQKOSQRTIDSLQIQKDEDYK-----1090
QY 758 SGKWNKESITKIDISSKYEYIFGNPKENKIIVKSKYLHELSTLQEI--RNNANSIDILE 816
DB 1091 ----SRYESKIEEAREKVTLLKGVKESQEIQYKSEIKKAEKELASSNTKITQLDQS 1145
QY 817 KKVMLTECEINVASNDROIVGRIEAEKNLSDSINYIKNEFKLIES-----ISDSLVDL 872
DB 1146 TKEL-----SSLKSNESKVAE--IESAKQQLSKKI-----EYSSIEADYNLQTELKQM 1194
QY 873 KHQGLDDSHFISFIDISTENGFRF-----RFINKETGNSIFETKE-----IFSEYA 922
DB 1195 KQTN-----TEVSTRITELSTKLNQVTEBAKSEIEKENQPPNPFMEEF 1239
QY 923 TH-----ISKEISNIKTIFDN-----VNGKLV-----KKVNL---DAHE 955
DB 1240 THMKLVNEQTASLRKEKPEFKLSBELQMLKERIMNGSLTSMDLTPKRSIAIGDKSMI 1299
QY 956 VNTLSAFTIQLSIEYNTTKESL-----SNLSVAMKVQVVAQLFSTGLNTI 1001
DB 1300 TWTVD-----FNKEIENLKFOLOQEQGNFQBAENYAIELQKLNKLTITGLNTN 1350
QY 1002 TDASKVVELVSTALDETIDLLPTLSEGLPIIATIDGVSGLGAARIKELSETNDPLLR-QBI 1060
DB 1351 TDYEK-----KYNDSQKRIITQL 1367
QY 1061 EAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELIQDKATKV 1120
DB 1368 ETKIGRLLAN-----SGG-----1380
QY 1121 IDYFKHISLAETGAFTLDDKIIMPQDDLVLSIEDFNNSITLKGCEIWAEGSGHTL 1180
DB 1381 -DYEDH-----HLDD-----NDD-----DWESRSSNGG-----SGGSTS 1409
QY 1181 TDDIDHFFSPSIT-----YRKPLWSIYDVLNKKKIDFSK-DLMVL-----1222
DB 1410 SSARNSLVKSESITAFAHMRGVSDYIYQIDITTKLSTREELNGSKTEILRLKALLRE 1469
QY 1223 -----PNAPNRVFGEMGWTGFRSLNDGTGLLDRIDRHYE-----GQFW 1264
DB 1470 SEDELYQVKQENYKTSVHDYEQ-----DLAQLKVKHETLLSRNKDINESLEIYKGRSDEY 1525

QY 1265 RYFAPADIATIKLPRIYEDTNVRINDGNTRSFIV--PVITTEQIRKNLSYFYSGSGS 1322
DB 1526 KKLLEAESAI--ATSKRHEEQATKEMKESRQLLLVREELATTQILLKDFR-----1574
QY 1323 YSLSLSPYNNMIDLNLVENDTVIDVNVVKNVITIESDEIQKG-ELIENILSKLNIEDNK 1381
DB 1575 -----IKVENLEATIEBKHHOLDANKKEIKQIQDKLNY-----1607
QY 1382 IILNHNHTINFG-DINES-----NRFISLTSILEDINIIIEIDLVSXSKYKILLSGN 1432
DB 1608 -----HLKNFENKELNKEELKLNLRDL-----DFKTDIETKLKENKQLQLDYE 1654
QY 1433 CMKLIENS-----SDIOQKIDHIGFNG---EHQKVIFYSYIDNETKNGFIDY 1477
DB 1655 DVLLVKNLQNEVEBELILOEELQNKIDELTNNRQLENEKLINEKRVNCTK-----1707
QY 1478 SKKGLFTAESNE-SIIRNIYMPDS-----NNLFYSSKOLKDIRIINKGDVKL 1526
DB 1708 -QISGL--KELVDEISIERDKLLKDKETLQNDLQMTNKFDTATTTELKQAH---GELDF 1760
QY 1527 L---IGNYFKD-----DMKVSLSFTIEDTNTIKLNGVYL---DEN-GVAQILKPMN 1570
DB 1761 LKKHLENQREDESAIKTELNOSKMTSPDIRDQ--KLNNELLVTKEENFSLVTKNKSIN 1818
QY 1571 NAKSALNT---SNSLMNPLES-----INIKNIFYNNLDPNTEFILDNTFIISG 1615
DB 1819 LKVSLEELKYSNEQLKYWESKVDTLKALOGALNEK---HEADTKIKNL-----1865
QY 1616 SNSIQGFELICDKNIOPIYF--INF-----KIKETSYTLVYGNRQNL 1656
DB 1866 QRSIKQLBIRVENESQSKRYNDENFDYQNKINHYKSTIDIHNEI 1912

RESULT 7

US-10-155-533-9
; Sequence 9, Application US/10155533
; Publication No. US20030104003A1
; GENERAL INFORMATION:
; APPLICANT: Nguyen, Thanh V.
; TITLE OF INVENTION: A No. US20030104003A1 Surface Protein of the Malaria Parasite
; FILE REFERENCE: 48417/CAB/R2682
; CURRENT APPLICATION NUMBER: US/10/155.533
; CURRENT FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 9
; LENGTH: 1610
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-155-533-9

Query Match 3.3%; Score 286.5; DB 15; Length 1610;
Best Local Similarity 18.2%; Pred. No. 5.1e-09;
Matches 335; Conservative 273; Mismatches 572; Indels 659; Gaps 93;

QY 15 VKFRI-----QEDRYVAILNALE-EYHNMSESVV-----EKY-----46
DB 79 LKYKINFCRNLYFNKNKNEYNICRGNKLNFNHIQTDNTIYKPKKKYEVGVKREKIKMYT 138
QY 47 -----LKLKOINLNTDNLVNTYKSGRKNALKKFEKYLTVLEVLKNNSLTPVEKN 97
DB 139 LPKDDKINTLKNBEESVTSITINKHDSVKVGRKKNTITSNENIMPNKSVSSINTN 198
QY 98 LHFTWIGQIINDTAINYINQKDVN-----SDYTVKFVYDSNAFLINTLTKKTIVESATN 151
DB 199 L-----SDSNKNDNSLNSKNDNTLNSKKNDNTLNSKKNDNTLNSKNNNSKNN 253
QY 152 NTLSPREN--LNDPEPDYKFKRMEIIVDKQHGFIDYKQSQIEEPEFIIDNIITY 209
DB 254 NSLNSKNSNSVNDKHID-----HIPEGKNKINN-----ID--VKHN 290
QY 210 LSNEYSKLEALNKYIESLNKITANNNDIRNLEKFADEDLVRLYNOELVERWNLAAS 269

Db 792 SLNE-----KFLVTANSIGIKARTKESSETISGPDQOELQALK--GNTSESTLKLQKE 844
Qy 160 NLNDPEFDYK-----FYKRMEL-----IYDKQHF--IDYKQIEENPEF 200
Db 845 KJOSTQAAKKLEGGINNMTROLFHUKSKSBAETOIKQREFFKMLTYEFFENTKRYDEL 904
Qy 201 IIDNIIKTYLSNEYSKDLBALANKYIBESLNKITTANNGDIRNL-EKFADEDLRLYNQEL 259
Db 905 QINNLANKS--NNEFKQINELSKIE---SLTEDNKNFAKQLEELKRLDTE---ENNEHL 955
Qy 260 VERWNLAAASDLIRISMLKEDGGVYLDVILPGIQDLPFKSINKPDSITNTSWEMIKLEA 319
Db 956 MDKLRASVA-----YNDLKAKSESE-----EE 979
Qy 320 IMKYKEYIPGYTSKNFMDLDEEV-----QRSFESALSXSCK-SEIFLPLDD-----IKV 368
Db 980 TVKAKEBELTTSK-IDNLEKELKEQQSKNLEGOLQONITDSTNEKPKLEDELKSIKK 1038
Qy 369 SPLVKIAPANNVINOALISLKSYSCLVINOIKRYKILNDNLNPSINEGTDFNTM 428
Db 1039 SNKE--ISSQSELIOKLEKTEKLOAKDBEIDKLAETKSNIDNLNSEI---SSLQSKL 1093
Qy 429 KIFSDKLASISNEDNMFMKITYNLKVGFAFPDVRSTINLSGPGVYTGAYQDLLMFKONS 488
Db 1094 KEAESHSSTKDEHSSL-----S 1111
Qy 489 TMIHLEPELRNPEFPKTKISOLTE-----QEITSLMFSNQARAKS 529
Db 1112 ENLKKLKEBYN-----TKTSMIAKLSAKIEBHKATDIEITKTHITDLOE-EHAKQKS 1165
Qy 530 QEEYKKGYPFEGALGDDNLDPAQNTVLDKDVYSKILSSMKTRNKYTHYVQLQGDKI 589
Db 1166 QFESERNDI-----KSNLDEANKELSD---NREKLSNLE-KETELNKKLTQEEKI 1213
Qy 590 SYEASCNLSPKOPYSILYQKNEGSETAYYYVADAIEKIDKRYPIYOISKNRNKILUT 649
Db 1214 SDETSVAISEDKSKSLKH-----DIEDLKREKIKLETTKENEETM 1255
Qy 650 FIGHGSEFNTDFANLD-----VDLSLSEIETI---LNLAKADISPKYIENLLGCNM 700
Db 1256 FKKEQLOVNDCKKELEACLLKLTETKEKENDLIRKLEAAKSD-----S 1300
Qy 701 FSYSIYAEBETYPGKLLKIKDRVSELMPISODSITVSANOVYERINEREGREILDHSGK 760
Db 1301 -----HTEKKSLLIEDTKSE-----SEKNVILNEQIEKLGEREKEVRDIQSO 1347
Qy 761 -----W-----INKEPSIHKDISKEYISFNPKENKIIVK-----S 791
Db 1348 LAAKTTDWEKIKITLTDKVLKESDLEK-TNKESVDTLKKEVENLKEISLLEDQKDDTT 1406
Qy 792 KYLHELSTLLOIRNANSDDIDLEKVMLTCEINVASN-----IDRQIVE 838
Db 1407 KY-KELAAQLETKSNLSDTTMELEK-----TELELKKVNELTTEATSELTKLQDNQSLT 1461
Qy 839 GRIEEA-----KNLTSDSINVINEPKLIBS-ISDSLYDLKHONGLDD 880
Db 1462 EBEIKTKAALTSSKDLVCGNQKSELQDSLSKVSSELKNPKENKYNQETSLADEBEKQ 1521
Qy 881 SHFIFGEDIKTENGFRIFINKTGNISFIEETEKEIFSEYATHI-----S 926
Db 1522 KEIVTLQ-----TELKDRISEVEKE--RAMLSENSETVIKEYSDKIKLESKINSIKENHS 1575
Qy 927 KEI-----SNIKOTI-----FDNVNGKLVKKYN-----LDAAEHVNTLNSAFPIQSIL 969
Db 1576 KEITTHNEOKTSLKODIAKLSQDHESAQOEDKENQNLKELKASLEKHNTESA---TSIE 1632
Qy 970 EYNTTKESLNSLVAMKVOVYLAQFSTGLNTITDASKVVVELYSTALDETIDLLPTLSEGL 1029
Db 1633 EKXNQIKELSEIYKSLKTEL-----KTSGDALKQSQKEYKTLTKTNSDTSFESKLEQLSEL 1687
Qy 1030 PIATIIDGVSIGAAIKELSETNDPLLRQIEIPAKIGIMAVNITAASTAIVTSGALGIASGF 1089

Db 1688 -----EKVKSDLQATDEKLGITEREIALKSELETNVKNSGLSTTS----- 1727
Qy 1090 SILLVPLAGISAGIPSLVNNLILQ-----DRATKVVDYF-KHISLAETEGFTLLDDKII 1144
Db 1728 -----ELAALTKTVKSLEKEEEOFLSGNKSKELEDYIQKHSIDISEKKAUT----- 1775
Qy 1145 MPQDDVLSEIDFNNSITLKGCEIWRAGGSGHTLTDIDHFFSSPSITTYRKPWLSIYD 1204
Db 1776 ---DELAKETKQFDDSKKKLLELE-----NDLTS-----TKKLETEKTQTSKFK 1817
Qy 1205 VLNIKKEK--IDFSKDLMLVLPNAPNRVFGYEMGWTGPRSLDNDGTK 1249
Db 1818 NLEERKDEIVKLNKELELLKN-----DNSGAK 1845

RESULT 9

US-09-801-368-56
; Sequence 56, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801.368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 2167
; TYPE: PRN
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-56

Query Match 3.1%; Score 268.5; DB 10; Length 2167;
Best Local Similarity 18.3%; Pred. No. 1.1e-07;
Matches 341; Conservative 276; Mismatches 610; Indels 633; Gaps 83;
Qy 53 NNLTDNYLNTYKSGRNKALKKFKLEYLTMEVLELKNLSLTP-----VEKNLHFIW 102
Db 141 SNAQDETPIAYKKSAAKSD-----SLLPSSSSLSPPQSRCSGTGLEKSLNTSG 190
Qy 103 I-----GGQINDTAINYNQWQDVNSDYTVKFFVYDSNAFLINTLTKTIVES-ATN----- 151
Db 191 ISNSGSGTNNNSNNNDNEQQRNVHLSNENYDITTVFKTGWNKSHGQTVATNYSMT 250
Qy 152 ---NTLESFRENLDNPEFDYN---KFY-----RKMELIY 180
Db 251 APSSSSSSSSQNLRNDAYSRRNSRFYNGDSSSLKNDSSSTTATNSGNDVASAKSSHAI 310
Qy 181 DKQKHFDY--YKSQIEENPEFIIDNIITKTYLSNEYSKDLBALANKYIEESLNKITTANNGN 238
Db 311 DPQMLVPDYRLYRAQ-----LKGCVNLNLYKSLNSNIKFFDPPTL---PASNSS 355
Qy 239 DIRNLEKFADEDLRLYNQELVERWNLAAASDLIRISMLKEDGG-----VYLDVDILPGI 293
Db 356 -----IANEN-----HQKKQQTNNQAOEALH---QKQSFQMGEPITLIDLKYLSEV 400

[illegible][illegible]

RESULT 10

US-10-032-585-7538
; Sequence 7538, Application US/10032585

```

: Publication No. US20030180953A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Terry, Roemer D.
: APPLICANT: Bo, Jiang
: APPLICANT: Charles, Boone
: APPLICANT: Howard, Bussey
:
: TITLE OF INVENTION: Gene Disruption Methods
:
: FILE REFERENCE: 01182-005-999
:
: CURRENT APPLICATION NUMBER: US/10/032,585
:
: CURRENT FILING DATE: 2001-12-20
:
: NUMBER OF SEQ ID NOS: 8000
:
: SOFTWARE: Patent In version 3.11

```

Query Match	3.1%	Score 266:	DB 12:	Length 1847:
-------------	------	------------	--------	--------------

Query Match: 3.14, SCORE 200, DB 12
Best Local Similarity 19.4%: Pred. No. 1.2e-07;

BEST LOCAL SIMILARITY 15.44; FREQ. NO. 1.2E-07;
Matches 366; Conservative 282; Mismatches 612; Indels 628; Gaps 99;

QY 4 VNKAQLQKMWYVKFRIQEDEYVAIILNALEEYHNMS-----ESSVVEKYLKLDI 52

Db 73 LMQASIEITGVLSNELDLDE----LNTAELLYNASDLKYKKGTSIGDSARLAYLYLRAHYI 128

QY 53 NNLTNDNYLNTYKK-----SGRNKA-----LKKPFKEYLTMEVLELKNNSLTPEVKNLHPIWI 103

Db 129 LNIV-GYLVSHKRLDIIITNNQVLFNDNLIKSPSKIYT---LSGKLNMDMKQK-----V 178

Qy 104 GGOINDTA-INYINQKDWNSDYTVKPVYDSNAFLINTLKKTIVESATNNTLESFRENLN 162

```

Db      179 TGDINLLAFINCINYSR-----SQLFNAHELLQGVVFGGLADNYYESY-GTLN 224

```

QY 163 DBEDYNNFYKRMELIYDKQK-----HF-----IDYKKSQIEENPEFIIDNIITKYL 211

Db . 225 N-----YKSLVEFISKNISDEDFVTHFLPSTLQLFKLLQLGEESLVDQFYKT-IT 275

Qy	212	NEWSKOLEALNKYIEESLNKITANNNGNDIRNLSEKFADEDL--VRLYNQBELVERWNLAAS	269
Db	276	SSILKOEYA-----NNFKSEIDILSKSKLSGFEIYTSF-----	309
Qy	270	DILRISMLKEDGGVYLDVDLPGIQDPLFKSINKPDSITNTSWEIMKLEIMKEYKEYIPG	329
Db	310	-----IFL--TEFIPWC-----KQSSSRFAKYDF--KDDILKYMFLIS	344
Qy	330	YTSKNPMDLBEVQSPESALSSKSKSEIFLPLDDIKVSPLEVKIAFANNNSVINQALIS	389
Db	345	Y-----GWERLL--	352
Qy	390	LKDSYCSGLVINGIQKNRYKILNDNLNPSINEGTDFTNTMKIFSKL--ASISNEDNNMFM	447
Db	353	---SYCSETANAKTQQVYDSN-----MYDFRALLQKNFPRLTAPKAFHYPGNQ--	397
Qy	448	IKITNYLKVCFAPDVRSTINLSGPGVYTGAYQDLLMPKONSTINHILLEPRLNPEPPTK	507
Db	398	ELLNAVRCGY-ENVSKLIDIS-----FULTDPSLNETLVSPFPQSF-----	437
Qy	508	ISOLTEQIEITSLMSFNQARAKSQPEEYKKGYPFGALGEDDNLDPQONTVLDDKDYVSKIL	567
Db	438	-----PSVFISNAAVMTSLRUSEEDFVLSSLNSEDEEESEDSDED-----	481
Qy	568	SSMKTRNKYIHYIYVLOQDKISYEASCNLFSDKPYSSILYQKNIEGSETAYYYYVADAE	627
Db	482	---SSTPKNKE--KSTGLDLDKIAQRAELERF---YLAFAYTYN-NRPELCAFLWGNEQV	532
Qy	628	IKEDIKRPIYQISNKEN--IKLTF-----IGHCKSEPT--DT	662
Db	533	THDIIGF-ITSGLANNTSPITATFCLLLGSLASAGAEATSRITWEILVHNNNASTRND	591
Qy	663	FANLDVDSLSEIETILNALKADISPKYIEINLLGCMFMSYSIVAEETYPGKLLLIKDR	722
Db	592	FSKISVDLSLSK-----YYID-----SLNESFEODLNAQLMLNQKKQ	630
Qy	723	VSELMPSISQDSITVSA-NOYEVRIINE-----GKEIILDHSGKWINKEESI-KDISSK	775
Db	631	DLFSTTTTSKQDLDDSGENRIVIELAEDSLVFISGFIQLLSAIVKNLNTKNSRKEIKSV	690
Qy	776	EYISFNP-----KENKIIYKSKYLHELSTLLOEIRNNANSSIDL-----EKKVMLT	822
Db	691	VYTRFSPIIRKGFUKFDNLINGSRFLQ---VDASIQSTNNPKFIDLPNVFSVDSRSIILT	746
Qy	823	ECBINVASNDRQIVEGRIBEAKNLTSDSINYIKNE-FKLIES-ISDSLVDLKHQNLDD	880
Db	747	NLILTFLGDF-----VTNDSPIYRIEILRWLVDRWYQGLHSLP-----	785
Qy	881	SHTISPEDISKTENGFR---IRFINKETGNSIFIEETEKEIFSEYATHISKEISNIKDTIF	937
Db	786	-----ED--KKDDAFPHIKRKYISKN-----VPINQAFSTNLTHLS-QIGNF--TV-	827
Qy	938	DNVNGKLVKVKNLDAAEVNTLNSAPFIQSILIEWNTTKESLNSLVAMKVQVTAQI.FSTG	997
Db	828	-----LVKKLLTPYADS-----NEAFKYSL-LYPCDLGSGYRFNNOLGWPYIEFLMQN	876
Qy	998	L--NTIIDASK-----VVELVSTALDET-----IDLPLTISEGLP---IATIID	1037
Db	877	VFANSATIANKRRVNLQMLLELFLSNALQEDVWKFLIDVAPKIIRDLKNFNGIFDSLIP	936
Qy	1038	GVSLG-----AAIKELSE--TNDPLLROEIEAKIGIMAVNLTAATAI.VTSALG	1084
Db	937	GVQLDPEFVKLHHSHVAVINYLFPENKTF.SAL.FKL---VNIGVDSVNESGESAAIVSHALG	993
Qy	1085	IAGFSILLVPLAGISAGIPSLVNNEILLODKATKV-----IDYFKHISLAETEGA	1135
Db	994	LIN--SLLRVQNSFINKLLP-----ILRNKDTQQOQHCCTAIGICTSMSLSLATPRTI	1044
Qy	1136	FTLLDDKIIIMPQD-----DLVL-----SEIDFNNNSITL-----G	1165
Db	1045	F----DCIYYPKNLGTGHVADFVEVILFHLISAVVQFALYVSCENTISKAISILKGVVSQ	1100
Qy	1166	KCEIWRAGSGSHTLTDD-----IDHPFSSPSIYTR--KPWLSIYDVLNKKGKIDFSKD	1218

Db	1101	KFFVTVSSADPLLNNDRLLTTFENIDESIKFAFDKFEEDSLNMKYILDF---	1157
Qy	1219	LMVLPNAPRVFGYMGWTPGPRSLNDGDKL---LDRIRDHYEQGYWRYFAFIADAL	1274
Db	1158	--VLGNL--NQPDGKVAATTAHFLL---GYKVGKDTLDLVQTN-----	1192
Qy	1275	ITKLKPRYEEDTVNRILNDGNWTRSFIVPVITTEQIRKNLSYFVGGSGSYSLSLSPYNNMI	1334
Db	1193	-----DQNTLLKSFNTLSISLDLI-----SEIDYNNGNHHIIDVGPAKLS-	1233
Qy	1335	DLNLVENDTWIIVDWNVKNIT---ESDEIQKGELIENILSKJAIENDKIILNNHTIN	1390
Db	1234	--SLILQILIKLCQDPISISITLQNLYEEL--PEKLVNCQPKLDL--NTVWCGN--Q	1284
Qy	1391	FYGD--INESNRFI-----SLTFSLIEDNIIIEI-----DLVSKSYKILLSGNC	1433
Db	1285	FDGDLQIDASNVPVDNQASTQAPFSPINORNLILQVLSLEFHSVKSRTKREYYSKVL	1344
Qy	1434	MKLIENSSDIQOKIDHIGF---NGEHQKIFYYSIDNE-----TKYNGFDYDYSK	1479
Db	1345	KEFVNRTPKVLTPLNTILNYSFRNFVQK---YEWLOQKFNMSLLLEAVNAQKNGTLDFS-	1400
Qy	1480	KEGLTAEFESNESIIRNIYMPDSNNLFI-----YS-SKOLDKIRI-----	1518
Db	1401	---VLTKVFLLCQTSNLITPESKQIFABEIMVEGSKISDFVTYKYSVSTDLDKDVQLKCLH	1457
Qy	1519	-----INKGDVYL-----LIGNFYKDDMKVS-----LSFTIEDTN	1548
Db	1458	SWCQLTEILVTDGGINSSNFLEVLQVIIPKINDYFDVDISFSEEMVSLCVLLFDLYDQS	1517
Qy	1549	TIKLVGYLDN-----GVAQILKFMWNAKSALNTSNLSMNFLESIN-IKNIIFYNNLDP	1601
Db	1518	TL-----ADRGEDFALGIERLIPLFOTCIAGILNSNSTPFSRSLDYVVGNGKFLCLKCFE	1571
Qy	1602	NIEFILDTFIIISGNSIGO--FELICD 1627	
Db	1572	RESFLQVNHII---KSVDDKKFFQVTCN 1596	

```

RESULT 11
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: EPITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 6281
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996

```

```

Query Match      3.0%; Score 257; DB 9; Length 6281;
Best Local Similarity 18.7%; Pred. No. 2.8e-06;
Matches 357; Conservative 291; Mismatches 708; Indels 556; Gaps 91;

```

```

QY 25 VAILNALEYHNMSSSVVEKYLKLDINNLTDNYINTYKSGRNKALKKFEYITMEVL 84
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3735 VTTMQAKDALNGDEKLAQKQEAALNLTLD--LNQPDALRNQINQAALATVEQT 3792
QY 85 ELKNSLTPVENLHFIWIGGQINDTAINYINQWKNVSDYTVKFVYDSNAFLNLTCKT 144
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3793 KQNAQNVNTAMSNLK---QGIANKDTVKASENYHDADAKQTAY---TNV---SQAE 3842
QY 145 IVESATNTNLE-----SFRENLD-----PEFDYKFKRMETIYDKQKHP 187
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3843 LINTQNTPTLDEITRALQTVTDKNGLNGEAKLATEKQNAKDAVSGMTHLNDKQAL 3902
QY 188 DYKQGIENPFIINIIKTYSNEYSKDLBALNKYIBESLNKITYTANNNDIRNLEKPA 247
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3903 ---KGQIDQSPETATVQVK-QTATSLDQAMBQLSQAINDKQA--TLADGNVNL---A 3951
QY 248 DEDLVRLYNQELVERWNLAASDILRLSMLEKDG--GVYLDVDILPGIOPDLFKSINKPD 305
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3952 DPKQNAVYKQA-----VAKAEALN-----KOSGTNEVQAVESITNEVNAKQALNGND 4001
QY 306 SITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVORSFESALSSK---SDKSEIFLP 362
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4002 NLAN-----AKQQAQQL-----ANLTHLNDKQKSFESITQAPLVDVTI--- 4044
QY 363 LDDIKVSPLEVKIAPANNSVI-NQALISLSDSCYSLVINQIKRY-----KILNDN 413
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4045 --NQAQTLDHAMELLRNSVADNQTTLASEDYHDATA---QRQNDYNOAVTAANNIQT 4099
QY 414 LNPSINEGDFNTTKMIFSDKLASISNEDNMFP-----MIKITYNLKVGPAVDV 462
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4100 TSPTWNPDDVNGATTQVNNTKVA-LDGDENLAAKQQAANNRLDQDLHLNNAKQQLSQSI 4158
QY 463 RSTINLSPGVVTGAYQDILLMFKDNSTNI---HLLPELNRNPEFPKTKISQTEOITS 519
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4159 TQSSDIAVNGCHKQTAESLNTAMGNLINAIAHQAVEQGNP-----INADTKQTAYN 4212
QY 520 WSNQARA-----KSQFEE--YKKGYPEGALGEDDNLDEFA----- 552
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4213 TAVNEAAMINKQTGNNAQTEVEQAITKVQTTLQALNGDHLNQVAKTNATQAIIDAL 4272
QY 553 ---QNTVLDKQVVS-----KKILSMKTRNKYIHYIVQLOGDKISYEASCNLFSD 601
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4273 NDPOKTAL-KDQVTAATLVTHQIEQNANTLN-QAMHGLROSIQDNAATKANSKYINED 4330
QY 602 PYSSILYOKNIEGSETAYVYVADABEIKEDKRYIPYQISKNENIKLTFIGHGKSEFND 661
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4331 QPEQNYDQAVQANN-----LINEQTATLD-----NNAINQAATVNTTKAALHGD 4377
QY 662 TPAFLNDVDSLSSEIETILNLAKADISPKYIEINLLGCNMFYSIVAEETYPGKLLIK 721
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4378 VKLQNDKDHAK---QTVSQLAHNLNAQKHEDTLIDSETTRTAVKQDLT-EAQLDOLMD 4433
QY 722 RVSELMPSISQDSITVSAQNOYVRIN-BEGKREILDHSGKWNKESIIKDISKEYISF 780
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4434 ---ALQQSIADKDATRASAY---VNAEPNKKQSYDEA---VQNAESIIAGLN----- 4477
QY 781 NPKENK-----IIVKSK-----YLHELSTLQOR-IRNNANS 810
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4478 NPTIINGNVSSATQAVISSKNALDGVRLAQDKQTAGNSLNHLDLQTLPAQQALLENQINN 4537
QY 811 SDIDLEKKVMLTECE----- 825

```

RESULT 12

US-09-932-183A-2

; Sequence 2, Application US/09932183A

; Patent No. US20020127641A1

; GENERAL INFORMATION:

; APPLICANT: Estell, David A.

```

DB 4538 ATTRGEVAQKLTEAQAALNQAMEALRNSIQDQOQTEAGSKFINEDPKQDAYQAAVQNAKD 4597
QY 826 -INVASN--IDRQIVEGRIBEAKNLTSDSINYIKNEFKLIESISLSYDLKHQGLDDSH 882
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4598 LINTQNTPTLTKAQQE-QLTQAVNAQKNLH---GDQKLADDKQHAVTDLNQLNGLNPNQ 4653
QY 883 -----FISFEDISKTENGFRIFINK----- 903
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4654 RQALSSQINNAATRGEVAQKLAELAKALDAQMALRNSIQDQOQTESG--SKFINEDPKQK 4711
QY 904 -----ETGNSPIFETEKEIFSYATHISKEISNIKDTTFIDVNGVKLVKV 948
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4712 DAYQAAVQNAKDLINQNTPTLTKSQVEQLQAVT-----TAKNHLGD--QKL 4758
QY 949 NLDAHAHEVNTLNSAFFIQLSIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKV 1008
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4759 ARDQQAQVTTVN-----ALPMLNHAQ-----QALTDAINAAPTREVA 4797
QY 1009 ELVSTALDETIDLLPTLSEGLPIIATIIDGVSIGAAIKELSETNDPDLRQETIEAKIGIMA 1068
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4798 QHVQTATE-----LDHMETLKNKVDQVNTDKA-----Q 4826
QY 1069 VNLTAAST---AIVTSALGIASGFSILLVPLAGISAGIPS-----LVNNEL 1111
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4827 PNYTEASTDKKEAVDQALQAAS---ITDPTNGSNANKDAVDQVLTQLQEKENELNGNER 4883
QY 1112 ILQDK--ATKVIDYFKHISLAETEGFTLLDDKI--IMPQDDLVLSIEDFNNSITLGKC- 1167
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4884 VAEATQAKQTTIDQTHLNADQIATAKONIDQATKLOPIAELVDQATQLNQSMQQLQAV 4943
QY 1168 -EIMRAEGSGHLLTDDIDHFPSSPSITVRKPWLSIYDVL--NIKEKIDFKDMLVLPN 1224
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4944 NEHANVEQTVDTQAD-----SDQONAYKQAIADAENVLKQNAKQVVD--QALQNILN 4995
QY 1225 APNRVFGYE---MGWTPGFRSLNDGTKLLDRIDHYEGQFYVRYFAFTADALITKLKPR 1281
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4996 AKQALNGDERVALAKTNGKHIDQ-----LNALNNAQODGF-----KGR 5034
QY 1282 YEDTVNRINLDCNTRSFTV-PVITTEQIRKNLSYSFYGGSGYSLSLSPYNNIDNLNVE 1340
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5035 IDQSN-----DLNQIQQVDEAKALNRAMDQLSQEITDNEGTGKSTNTVYNADTQVKQY 5089
QY 1341 NDTWIDVD-----NVVKNTIIESDEIQKGLIENILSKLNTEDNKIILNN----- 1386
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5090 DET-----VDKAKQALDKSTGQNLTK-QVIKLNDVATAAKKALNGEER---LNNRKAEL 5141
QY 1387 HTINPYGDIENSRFISITPFILEDINIIIEIDLVSYSKYLILSGNCKLIENSIDIQK 1446
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5142 QRLDQLTHLNAQROLA-----IQQIN---NAETLNKASRAI--NRATKLDNAMGAVQY 5191
QY 1447 ID--HIGFNGEHQKIFYYSYI---DN-ETKYNQFI-----DYSKKEGLFTAESFESIIR 1495
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5192 IDEQHLGVISS-----TNYINADNLKANYDNATANAHAHELDKVGQNAIAKAEALQK 5245
QY 1496 NIY-----MPDSNNLFIYSKOL-----KDI--RIINKGDVKLLIGNYFKDD 1535
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5246 NIIDAQNALNGDONLANAKDKANAFVNSLNGLNQOQDLAHKAINNAD----- 5293
QY 1536 MKVSLSTIEDTNTIKLANGVYLDENGVAQILKFM-----NNAKSALN-----TSNSLMNPL 1586
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5294 -----TVSDVTDIVNN--QIDLDNDAEMETLKHLDVNEIPNAEQTVNYQADNNAKTNPD 5344
QY 1587 ESINTKNFYNNLDNPIEFILDTNFIISGNSI-----GQFELICDKDKNIQ 1633
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5345 DAKRLANTLLNSDNTNVN---DINGAIQAVNDIAHNLNGDQRLQDAKDKAIQ 5393

```

```
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-09-932-183A-2

Query Match      3.0%; Score 256; DB 10; Length 2285;
Best Local Similarity 17.8%; Pred. No. 7e-07;
Matches 313; Conservative 279; Mismatches 595; Indels 568; Gaps 76;

QY 142 KKTIVSATNN-----TLSEFREN---LNDPEFYNNKFKYKMEII-----Y 180
DB 142 KKTIVSATNN-----TLSEFREN---LNDPEFYNNKFKYKMEII-----Y 180
QY 181 DKQKHFDYKSOIENPEFIIDNIIKTVLSNEYSKDLALNKYTESLNKIKITANNNDI 240
DB 181 DKQKHFDYKSOIENPEFIIDNIIKTVLSNEYSKDLALNKYTESLNKIKITANNNDI 240
QY 841 DSGGNAILTKNKELEK---AIEN-TKEYLALKQETRSACKTFEDASKEIKSK-DEL 894
DB 841 DSGGNAILTKNKELEK---AIEN-TKEYLALKQETRSACKTFEDASKEIKSK-DEL 894
QY 241 RNLEKFADEDLVRLNQEOLVERNLAA-----ASDILRISMLK-----EDG----- 281
DB 241 RNLEKFADEDLVRLNQEOLVERNLAA-----ASDILRISMLK-----EDG----- 281
QY 895 KQYKQIAD-----YNDKGRPKWDLIADDDYKVAADKAKQSMKLAQSDIESGNKVKDS 948
DB 895 KQYKQIAD-----YNDKGRPKWDLIADDDYKVAADKAKQSMKLAQSDIESGNKVKDS 948
QY 282 -----GVYLDVILPGIQPLFKSINK-----PDSITNTSWEMIKLEAIMKYKEYTP 328
DB 282 -----GVYLDVILPGIQPLFKSINK-----PDSITNTSWEMIKLEAIMKYKEYTP 328
QY 949 VLSIANAYSIDISNTLTKTSISDVNKNLKDLDLPEELEKFSLSGLKQE--KMQKALD 1006
DB 949 VLSIANAYSIDISNTLTKTSISDVNKNLKDLDLPEELEKFSLSGLKQE--KMQKALD 1006
QY 329 GYTSKNFMDLBEVQSFESALSSKSKSEIFLPLDDIKVSPLEVKIAPNNSVINQALI 388
DB 329 GYTSKNFMDLBEVQSFESALSSKSKSEIFLPLDDIKVSPLEVKIAPNNSVINQALI 388
QY 1007 SGDEKAFDNKADQLSLLET--YKSDSS-----IDVFKMS-----FDKAK 1046
DB 1007 SGDEKAFDNKADQLSLLET--YKSDSS-----IDVFKMS-----FDKAK 1046
QY 389 SLKSDYCSDLVINQIKRYKILNDNLNPSINEGTDFNTMTKIPSDKLASISNEDNMFM 448
DB 389 SLKSDYCSDLVINQIKRYKILNDNLNPSINEGTDFNTMTKIPSDKLASISNEDNMFM 448
QY 1047 NIKD-----GDKSLSVKSEVCDLGETLAEGNEAEDEGK----- 1081
DB 1047 NIKD-----GDKSLSVKSEVCDLGETLAEGNEAEDEGK----- 1081
QY 449 KITNYLKVGFAPDVRSTINLSGGVYTGAYQDILL---MFKDNTNTHLLEPELRNPEFPK 505
DB 449 KITNYLKVGFAPDVRSTINLSGGVYTGAYQDILL---MFKDNTNTHLLEPELRNPEFPK 505
QY 1082 KLKEALDANSVDIIKAAIKEMSDAMQFDSVQDVLNGDIFNNTKQDVAPLNDLLEKMAEGK 1141
DB 1082 KLKEALDANSVDIIKAAIKEMSDAMQFDSVQDVLNGDIFNNTKQDVAPLNDLLEKMAEGK 1141
QY 506 TKISQLTEQEITSLWSPNQARAKSQPEEYKGYFEGALGEDDNLDFQNTVLDKDYVSK 565
DB 506 TKISQLTEQEITSLWSPNQARAKSQPEEYKGYFEGALGEDDNLDFQNTVLDKDYVSK 565
QY 1142 S-----ISANE-----NTLIQKQ-----KE 1157
DB 1142 S-----ISANE-----NTLIQKQ-----KE 1157
QY 566 ILSSMKTRKEYIHYIYVOLQGDKI SYEASCNLPFKDPYSSILYKQNEGSETAYVYVAD 625
DB 566 ILSSMKTRKEYIHYIYVOLQGDKI SYEASCNLPFKDPYSSILYKQNEGSETAYVYVAD 625
QY 1158 LAQAIENG-----VKNRDE-----VIQRKV----- 1182
DB 1158 LAQAIENG-----VKNRDE-----VIQRKV----- 1182
QY 626 AEIKEDKIRIPYQISKNRNIKLTFIGHGKSEFNTDTFANLDVDSLS----- 673
DB 626 AEIKEDKIRIPYQISKNRNIKLTFIGHGKSEFNTDTFANLDVDSLS----- 673
QY 1183 -----KLDAYNDMVTYSNK--LMKTEVNNAIKTLNADT---LRIDSLLKLRKERKLDMSA 1233
DB 1183 -----KLDAYNDMVTYSNK--LMKTEVNNAIKTLNADT---LRIDSLLKLRKERKLDMSA 1233
QY 674 -----EIEITLNLAKADISPKYTEINLGCNMFPSIYABETYPGKLLKIKDRVSELM 728
DB 674 -----EIEITLNLAKADISPKYTEINLGCNMFPSIYABETYPGKLLKIKDRVSELM 728
QY 1234 ELSDLVKSINNADAKELKLEEMKLPQGGYSNQI-----BAMQSVKSALESYIS 1286
DB 1234 ELSDLVKSINNADAKELKLEEMKLPQGGYSNQI-----BAMQSVKSALESYIS 1286
QY 729 SISQDSITVTSANQ-----YEVNRNEEGK-----REILDHSGKWINK 764
DB 729 SISQDSITVTSANQ-----YEVNRNEEGK-----REILDHSGKWINK 764
QY 1287 ASEATSTQEMNKQALVEACTSLNWTDOQEKANEETKTSMYVVDKYKALEKVNAEIDK 1346
DB 1287 ASEATSTQEMNKQALVEACTSLNWTDOQEKANEETKTSMYVVDKYKALEKVNAEIDK 1346
QY 765 EESIIKDIS--SKYEISFNPKNKII--VKSKYLHELSTLLQEIIRNNANSDDILEKVM 821
DB 765 EESIIKDIS--SKYEISFNPKNKII--VKSKYLHELSTLLQEIIRNNANSDDILEKVM 821
QY 1347 YNQVNDYPKYSQYRDAIKKEIKALQKKLMQEQAKLLKQIKSGNITQYGIIVTSTTS 1406
DB 1347 YNQVNDYPKYSQYRDAIKKEIKALQKKLMQEQAKLLKQIKSGNITQYGIIVTSTTS 1406
```

RESULT 13
US-09-726-949A-1
; Sequence 1, Application US/09726949A

	Query Match	2.9%	Score 249;	DB 9;	Length 1706;
	Best Local Similarity	19.5%;	Pred. No. 1.2e-06;		
	Matches 349;	Conservative 261;	Mismatches 650;	Indels 534;	Gaps 90;
QY	95	EKNLHFIIWIGGQNDTA	INYNOKKVDNSDYVKFVYDSNAFLNTLTKKTVESATNTTL	154	
Db	111	ESNLRF-----NSSY	LOGTNQITGRYEDGTLSTSTSD-----	157	
QY	155	ESPRENL-NDPEFDY	NKF-YRKRMELIYDKQKHFIDYKYSQIEENPEFIIDNI	212	

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:27:53 ; Search time 23.75 Seconds
(without alignments)
3028.567 Million cell updates/sec

Title: US-09-126-816B-6_COPY_1_1700
Perfect score: 8677
Sequence: 1 MNLVKAQLQKVVYKPRIQ.....YLYGIDRVYVNVIIAPNLYT 1700

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6733.5	77.6	2366	1 US-08-480-604A-10	Sequence 10, Appl
2	6733.5	77.6	2366	2 US-08-405-496A-10	Sequence 10, Appl
3	6733.5	77.6	2366	3 US-08-915-136-10	Sequence 10, Appl
4	6733.5	77.6	2366	4 US-08-957-310-10	Sequence 10, Appl
5	6733.5	77.6	2366	4 US-10-011-366-10	Sequence 10, Appl
6	4458	51.4	2710	1 US-08-480-604A-6	Sequence 6, Appl
7	4458	51.4	2710	2 US-08-405-496A-6	Sequence 6, Appl
8	4458	51.4	2710	3 US-08-915-136-6	Sequence 6, Appl
9	4458	51.4	2710	4 US-08-957-310-6	Sequence 6, Appl
10	4458	51.4	2710	4 US-10-011-366-6	Sequence 6, Appl
11	501.5	5.8	3169	4 US-09-453-702B-257	Sequence 257, App
12	348	4.0	10182	4 US-09-134-001C-3159	Sequence 3159, Ap
13	310	3.6	2184	4 US-09-417-485D-6	Sequence 6, Appl
14	287	3.3	3433	4 US-09-091-501B-10	Sequence 10, Appl
15	281	3.2	3696	4 US-09-134-001C-5080	Sequence 5080, Ap
16	266.5	3.1	2662	4 US-09-595-684B-31	Sequence 31, Appl
17	261.5	3.0	3248	1 US-08-353-700-1	Sequence 1, Appl
18	261.5	3.0	3248	5 PCT-US95-16216-1	Sequence 1, Appl
19	261	3.0	2504	1 US-09-328-352-5821	Sequence 5821, Ap
20	257.5	3.0	3135	1 US-08-323-170B-2	Sequence 2, Appl
21	257.5	3.0	3135	4 US-08-954-441-2	Sequence 2, Appl
22	256	3.0	3285	4 US-09-308-375-2	Sequence 2, Appl
23	253	2.9	1296	1 US-08-480-604A-28	Sequence 28, Appl
24	253	2.9	1296	2 US-08-405-496A-28	Sequence 28, Appl
25	253	2.9	1296	3 US-08-915-136-28	Sequence 28, Appl
26	239	2.9	2391	2 US-08-446-855A-2	Sequence 2, Appl
27	250	2.9	2391	3 US-09-150-741-2	Sequence 2, Appl

28	249	2.9	4536	4 US-09-180-422B-27	Sequence 27, Appl
29	249	2.9	4563	4 US-09-108-006C-1	Sequence 1, Appl
30	241	2.8	1786	3 US-08-973-462-8	Sequence 8, Appl
31	239.5	2.8	1169	4 US-09-255-829-20	Sequence 20, Appl
32	233.5	2.7	3878	4 US-09-914-259-11	Sequence 11, Appl
33	233	2.7	2522	3 US-09-251-645-13	Sequence 13, Appl
34	231.5	2.7	1346	1 US-08-471-033-23	Sequence 23, Appl
35	231.5	2.7	1346	2 US-08-471-044-23	Sequence 23, Appl
36	231.5	2.7	1346	2 US-08-463-483A-23	Sequence 23, Appl
37	231.5	2.7	1346	2 US-08-471-046A-23	Sequence 23, Appl
38	231.5	2.7	1346	2 US-08-470-566B-23	Sequence 23, Appl
39	231.5	2.7	1346	2 US-08-469-334-23	Sequence 23, Appl
40	231.5	2.7	1346	3 US-09-300-529-23	Sequence 23, Appl
41	230.5	2.7	3290	4 US-09-328-352-5486	Sequence 5486, Ap
42	226.5	2.6	1338	1 US-08-471-033-50	Sequence 50, Appl
43	226.5	2.6	1338	2 US-08-471-044-50	Sequence 50, Appl
44	226.5	2.6	1338	2 US-08-463-483A-50	Sequence 50, Appl
45	226.5	2.6	1338	2 US-08-471-046A-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-10
; Sequence 10, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

; REFERENCE/DOCKET NUMBER: OPHD-01763

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410

; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2366 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-480-604A-10

Query Match

Best Local Similarity 76.4%; Pred. No. 0;

Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;

QY 1 MNLVNAQAKQVYKFRIOEYVAILNALBEYHNMSSESVVEKYKLKDINNTDNYL 60
DQ 1 MSLVNRKQLEKMANVRFRTOEYVAILDALBEYHNMSSESVVEKYKLKDINSLTDIYI 60
QY 61 NTYKSGRNKALKKPKKYLTMVELEKNNSLTPVEKNLHFIMTGGQINDTAINYNQWD 120
DQ 61 DTYKSGRNKALKKPKKYLTMVELEKNNSLTPVEKNLHFIMTGGQINDTAINYNQWD 120
QY 121 VNSDYTVKPYDSNAPLINTLTKTIVESATNNTLESFRENLDNPEFDYKFKRMEIY 180
DQ 121 VNSDYTVKPYDSNAPLINTLTKTIVESATNNTLESFRENLDNPEFDYKFKRMEIY 180
QY 181 DQKHPIYKQIENPFIIDNIITKYLSEYKDLALNKYIIEESLNKTIANNNDI 240
DQ 181 DQKHPIYKQIENPFIIDNIITKYLSEYKDLALNKYIIEESLNKTIANNNDI 240
QY 241 RNLEKFADELRLYNQELVERNLAAASDIIRISMLKEDGGVLDVDTLPGIQDLPFS 300
DQ 241 RNLEKFADELRLYNQELVERNLAAASDIIRISMLKEDGGVLDVDTLPGIQDLPFS 300
QY 301 INKPSITNTSMEMIKLEAIMKYKEYIPGVTSKNFDMLEEVQRFESALSSKSKSEIF 360
DQ 301 INKPSITNTSMEMIKLEAIMKYKEYIPGVTSKNFDMLEEVQRFESALSSKSKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLKDYSQSDLVINQIKRYKILNLDNLPSE 420
DQ 361 LPDDIKVSPLEVKIAFANNVINOALISLKDYSQSDLVINQIKRYKILNLDNLPSE 420
QY 421 GTDFNTMTKIFSDKLASINEDNMEMIKITNLYKVPADYRSTINLSGPGVYTCAYD 480
DQ 421 GTDFNTMTKIFSDKLASINEDNMEMIKITNLYKVPADYRSTINLSGPGVYTCAYD 480
QY 481 LLMFKONSTNIHLLEPELRFNFPPTKISQLEQETISLWSFNQARAKSQFEYKGYPE 540
DQ 481 LLMFKONSTNIHLLEPELRFNFPPTKISQLEQETISLWSFNQARAKSQFEYKGYPE 540
QY 541 GALGEDNDLDFQNTVLDKDYVSKKILSSMKTRNKEYIHYIVOLQGDKISYEASCNLF 600
DQ 541 GALGEDNDLDFQNTVLDKDYVSKKILSSMKTRNKEYIHYIVOLQGDKISYEASCNLF 600
QY 601 DPVSSITLYOKNETGSTAYVYVADAEIKEDKIRIPYOISKNKMLKTFIGHGSEFNT 660
DQ 601 DPVSSITLYOKNETGSTAYVYVADAEIKEDKIRIPYOISKNKMLKTFIGHGSEFNT 660
QY 661 DTFANLDVDSLSSEIETILNLAKADISPKYIEINLGCNMFYSIYAEETYPGKLLKIK 720
DQ 661 DTFANLDVDSLSSEIETILNLAKADISPKYIEINLGCNMFYSIYAEETYPGKLLKIK 720
QY 721 DRVSELMPSISQDSITVSANQVEVRINEEGKREILDHSGKWNKESIIKDISKEYISF 780
DQ 721 DRVSELMPSISQDSITVSANQVEVRINEEGKREILDHSGKWNKESIIKDISKEYISF 780
QY 781 NPENKIIYKSKYLHELSTLLOEIRNNSDDIDLEKVMTECEINVASNIDROIVEGR 840
DQ 781 NPENKIIYKSKYLHELSTLLOEIRNNSDDIDLEKVMTECEINVASNIDROIVEGR 840
QY 841 IBEAKNLTSDSINYIKNEFKLIESDSLYDLKHQNGLDDSHFISPEDISKTEGFRIFR 900

Db 841 IEEAKNLTSDSINYIKNEFKLIESDSLYDLKHQNGLDDSHFISPEDISKTEGFRIFR 900
QY 901 INKETGNSIFETEKEIEFSEYATHISKEISINIKDTIFDNVNGKLVKKNLDAHEVNTLN 960
DQ 901 INKETGESIFVETEKITFSEYANHITBESIKTGITFDTVNGKLVKKNLDTTHEVNTLN 960
QY 961 SAFFIOSLIEYNTTKESLSNLSVAMKVQVYLAQFSTGLNTITDASKVELVSTALDETD 1020
DQ 961 SAFFIOSLIEYNTTKESLSNLSVAMKVQVYLAQFSTGLNTITDASKVELVSTALDETD 1020
QY 1021 LLPTLSEGLPIATIIDGVSLGAAIKELSETNDPLRQIEIAKIGIMAVNLTAASTAVT 1080
DQ 1021 LLPTLSEGLPIATIIDGVSLGAAIKELSETNDPLRQIEIAKIGIMAVNLTAASTAVT 1080
QY 1081 SALGIASGFSILLVPLAGISAGISLVNNELILQDKATKVIDYFKHISLAETEGAFLLD 1140
DQ 1081 SSGIASGFSILLVPLAGISAGISLVNNELILQDKATKVIDYFKHISLAETEGAFLLD 1140
QY 1141 DKIMPODDLVLSEIDFNNSITLTKCEIWRAGGSGHTLTDIDHFFSSPSITYRKPL 1200
DQ 1141 DKIMPODDLVLSEIDFNNSITLTKCEIWRAGGSGHTLTDIDHFFSSPSITYRKPL 1200
QY 1201 SIYDVNLTKKEDIDFSDKDLMLPNAPNRFVGYEMGTGFRSLDNDGTGKLLDIRDHVEG 1260
DQ 1201 SIYDVNLTKKEDIDFSDKDLMLPNAPNRFVGYEMGTGFRSLDNDGTGKLLDIRDHVEG 1260
QY 1261 QFYRYFAFIADALITKLPYEDTNVINLDGNTRSFIVPVITTEQIRKNLSYFYSGG 1320
DQ 1261 QFYRYFAFIADALITKLPYEDTNVINLDGNTRSFIVPVITTEQIRKNLSYFYSGG 1320
QY 1321 GYSLSLSPYNNIDNLVENDTVIDVDNVVKNITTESDEIOKGLTENILSKLNEDN 1380
DQ 1321 GYSLSLSPYNNIDNLVENDTVIDVDNVVKNITTESDEIOKGLTENILSKLNEDN 1380
QY 1381 KIILNNHTINFGDINESNRFIISLTFSILEDINIIEIDLVSYSKYILLSGNCKMLIENS 1440
DQ 1381 KIILNNHTINFGDINESNRFIISLTFSILEDINIIEIDLVSYSKYILLSGNCKMLIENS 1440
QY 1441 SDIQOQKHIFGNGHQKIFVYSYIDNETKNGFDIDYKKEGLFPAEBSNESIIRNIYWP 1500
DQ 1441 SDIQOQKHIFGNGHQKIFVYSYIDNETKNGFDIDYKKEGLFPAEBSNESIIRNIYWP 1500
QY 1501 DSNNFIYSKDLDIRIINKGDVKKLLIGNYFKODMKVLSFTIEDTNTIKLVYVLDEN 1560
DQ 1501 DSNNFIYSKDLDIRIINKGDVKKLLIGNYFKODMKVLSFTIEDTNTIKLVYVLDEN 1560
QY 1561 GVAQILKFMNNAKSALNTSNLMNPLESINIKNIFFNNLDNPIEFLDNTNFIISGNSIG 1620
DQ 1561 GVAQILKFMNNAKSALNTSNLMNPLESINIKNIFFNNLDNPIEFLDNTNFIISGNSIG 1620
QY 1621 QFELICDKXKNIQPYFINFKIKETSITLYVGNRQNLIVEPSYHLDSDSGNISSTVINFSOK 1680
DQ 1621 QFELICDKXKNIQPYFINFKIKETSITLYVGNRQNLIVEPSYHLDSDSGNISSTVINFSOK 1680
QY 1681 VLYGIDRVYVKNVIAPNLYT 1700
DQ 1681 VLYGIDRVYVKNVIAPNLYT 1700

RESULT 2

US-08-405-496A-10

; Sequence 10, Application US/08405496A

; Patent No. 5919685

; GENERAL INFORMATION:

; APPLICANT: WILLIAMS, JAMES A.

; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESS: MEDLEN & CARROLL, LLP

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94104
 COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/405,496A
 FILING DATE: 16-MAR-1995
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/329,154
 FILING DATE: 25-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/161,907
 FILING DATE: 02-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/985,321
 FILING DATE: 04-DEC-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/429,791
 FILING DATE: 31-OCT-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: INGOLIA, DIANE E.
 REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPHD-01308
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2366 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-405-496A-10

Query Match 77.6%; Score 6733.5; DB 2; Length 2366;
 Best Local Similarity 76.4%; Pred. No. 0;
 Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;

Qy	1	MNLVKAQIQKVVYKFRQDEYVAILNALBEEYHNMSESSVVEKYKLDKINLTDNYL	60
Db	1	MSLVNRKQLEKXANVRFRQDEYVAILDALBEEYHNMSENTVEKYKLDKINSLTDYI	60
Qy	61	NTYKSGRNKALKKPKFYLTMEVLELKNNSLTPVEKNLHFVIGQINDTAIYINQWKD	120
Db	61	DTYKSGRNKALKKPKFYLTVEVLELKNNSLTPVEKNLHFVIGQINDTAIYINQWKD	120
Qy	121	VNSDYTVKFDYDSNAPLINTLTKTTIVESATNNTLSEFRENLDNPDYDKFKRMETIY	180
Db	121	VNSDYVNVVFDYDSNAPLINTLTKTTIVESAINDTLSEFRENLDNPDYDKFKRMETIY	180
Qy	181	DKQKHFIDYKSOIENPFIIDNIKTYSNEYSKDLEALANKYTEESLUNKITANNNDI	240
Db	181	DKQKFIYNYKAQRENPELIIDDIKTYLSNEYSKDELNTYITEESLUNKITQNSGNDV	240
Qy	241	RNLKPADBDLVRNLQNELVERWNLAASDILRISMLKEDGGVLDVILPGIQDLPKFS	300
Db	241	RNFEFKNGESFNLYEQELVERWNLAASDILRISALKEIGMYLDVDMPLGIQDLPES	300
Qy	301	INKPSITNTSHEMIKLEAIMKYKEYIPGTYTSKNFMDLDEEVQSFESALSSKDKSEIF	360
Db	301	IEKPSVTVDFWEMTKLEAIMKYKEYIPEYTSHEFDMLDEEVQSFESVLSKDKSEIF	360
Qy	361	LPDLDDIKVSPLEKVFAPANNVINOALISLKDYSYCSDLVINOLKNRYKILNDNLNPSINE	420
Db	361	SSLGMEASPLEVKIAFNKGIINQGLISVKOSYCSNLIVKQIENRYKILNNSLNPASE	420
Qy	421	GTDFNTTKMIFGDKLASINEDNMFMKIKITNYLVKGFAPDVRSTRINLSGGVYTGAYOD	480

Db	421	DNDFNTTTNTFIDSIMAEANADNGRFMMELGKYLVRGFFPDVKTNTINLSGPEAYAAAYOD	480
Qy	481	LLMFKDNSTNIHLLEPELRNPEFPKTKISQLTQEITSLWSFNOARAKSQREYKKGYPE	540
Db	481	LLMFKEGWNTHLIEADLRNPEISKTNISQSTEQEWASLWSFDDARAKAQFBEYKKNYFE	540
Qy	541	GALGEDDNLDAQNTVLDKQVYVSKILSSMKTRNKEYIHYIVOLQGDKISYBASCNLFK	600
Db	541	GSLGEDDNLDFSQNIWVDKEYLEKISSLARSSERYIHYIVOLQGDKISYEAACNLFAK	600
Qy	601	DPYSSILYQKNIEGSETAYYYVADABEIKDYRIPYQISNKRNIKULTFIGHGSEFWT	660
Db	601	TPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNT	660
Qy	661	DTFANLDVDSLSSEIETILNLAKADISPKYIEINILGCNMFPSYSIVAEETYPCKLLKLIK	720
Db	661	DIFAGFDVDSLSSEIETIEAIDLAKEDISPKSIEINILGCNMFPSYSINVEETYPCKLLKLVK	720
Qy	721	DRVSELMPSISQDSITVSANQYEVRAINBEGKREILDHSGKWINKESIIKDISKEYISF	780
Db	721	DKISELMPSISQDSIIVSANQYEVRAINSEGRELLDHSGEWINKESIIKDISKEYISF	780
Qy	781	NPENKILVSKYLHELSTLLOEIRNANSSDIDLEKKVMLTECEINVASNIDQIIVER	840
Db	781	NPENKITVYKKNLPESLTLLOEIRNNSNSDIELEEKVMLTECEINVISNIDTQIIVER	840
Qy	841	IEEAKNLTSDSINYIKNEPKLIESDSLYDLKHONGLDSDHFI SFEDISKTENGFRIRP	900
Db	841	IEEAKNLTSDSINYIKDEFKLIESDLCALCDUQNELEDSDHFI SFEDISEDEGFSIRP	900
Qy	901	INKETGNSIFITEKEIFSEYATHISKEISNIKDTIFPNVNGKLVKKNLDAAEVNTLN	960
Db	901	INKETGESIFVETEKTFSEYANHITBELSKIGTIFDTVNGKLVKKNLDTTHEVNTLN	960
Qy	961	SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAOLFSTGLNTITDASKVVELVSTALDETID	1020
Db	961	RAFFIQSLIEYNSKESLSNLSVAMKVQVYAOLFSTGLNTITDAAKVVELVSTALDETID	1020
Qy	1021	LLPTLSEGLPIIATIIDGVSIGAAIKELSETNDPILROBIEAKIGIMAVNLTAASATVIT	1080
Db	1021	LLPTLSEGLPIIATIIDGVSIGAAIKELSETSDPLROBIEAKIGIMAVNLTTATTAIT	1080
Qy	1081	SALGIASGFSILLVPLAGISAGIPSLVNNELLDQKATKVIDYFKHISLAETEGFTLLD	1140
Db	1081	SSGIASGFSILLVPLAGISAGIPSLVNNELVLRKATKVVDYFKHVSIVETEGVFTLLD	1140
Qy	1141	DKIMPDQDLVLSIEDFNNSITLQKCEIWRABGSGGHTLTDIDHFPSSPSITVRKPL	1200
Db	1141	DKIMPDQDLVLSIEDFNNSIVLQKCEIWRMEGGSGHTVTDIDHFPSPSITVREPHL	1200
Qy	1201	STYDVLNTKKEKIDFSKDLVLPNAPNRVGVEMGWTGCFRSLDNDGTLLDRIDHVEG	1260
Db	1201	STYDVLVEQKSELDSKDLVLPNAPNRVFAWETGWTGCLSLDNDGTLLDRIDRNYEG	1260
Qy	1261	QPYRWYFAFIADALITLKPREDTNVNRINLDGNTRSPFVITTEQIRKNLSYFYGSG	1320
Db	1261	EYRWYFAFIADALITLKPREDTNIRINLDNTRSPFVITTEYREKLSYFYGSG	1320
Qy	1321	GSYSLSLSPYNNIDNLNVENDTWIDVNVVKNITIESEDEIQKGLIENILSKLNIEDN	1380
Db	1321	GYALSLSQYNNGINIELSESQVWIIDVNVVDRVTIESDKIKKGLIEGILSTLSIEN	1380
Qy	1381	KILNNHTINFGDINESNRPISTLFTSILEDINIIIEIDLVSYSKILLSGNCKLIENS	1440
Db	1381	KILNHSHEINFGVNGSNGFVSLTFSLLEGNAIIEVDLLSKYSKLLTSGSLKILMLNS	1440
Qy	1441	SDIQOKIDHIGFNGSHQYIFYSIDNETKXNGPDYDKKGLTAEAFENESGIIRNIYP	1500
Db	1441	NHIQOKIDYIGNSLQKNIPYSFVDSGKENGFGINGSTKEGLFVSELDPVVLISKVYND	1500
Qy	1501	DSNNLFYSSKDLKDIRIINKGDKVLLIGNYFKDDMKVSLSTFIETDNTIKLVGVYLDN	1560
Db	1501	DSKPSFGYVNNLKDVKITKONVNILTYGULKDDIKISLSLTQDEKTIKLSNVHLDGS	1560

QY 1561 GVAQILKFMNNAKSAALNTSLSMNFLESINIKNIFFVNNLDNPIEFILDTNPLISGNSIG 1620
Db 1561 GVAELIKFMR-KGNTNTSDLSMFLSNNIKSIFVNFLOSNIKFILDNFIISGTTISG 1619
QY 1621 QFELICDKDKNIQPIYFINKIKETSITLVGNRQNLIVEPSYHLDSDGNSISSTVINFSQK 1680
Db 1620 QFEFICDNDNIQPIYFINKIKETSITLVGNRQNLIVEPSYHLDSDGNSISSTVINFSQK 1679
QY 1681 YLYGIDRYNKVILIAPNLT 1700
Db 1680 YLYGIDSCVKNVISPNIYT 1699

RESULT 3

US-08-915-136-10
; Sequence 10, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-915-136-10

Query Match 77.6%; Score 6733.5; DB 3; Length 2366;
Best Local Similarity 76.4%; Pred. No. 0;
Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;
QY 1 MNLVNAQLOKQWVYKRIQDEYVAILNALLEEYHNHNSSESSVVEKYLKLDNNLTNYL 60
Db 1 MSLVNRKQLEKQANVRFTQDEYVAILDALLEEYHNHNSSENTVVEKYLKLDNLSUTDIYI 60
QY 61 NTYKSGRKNKALKKFKEYLTMVLELKNNSLTPVEKNLHFIIWIGQINDTAINYNQWKD 120
Db 61 DTYKSGRKNKALKKFKEYLTMVLELKNNSLTPVEKNLHFVWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKFDVDSNAFLINTLTKTIVESATNTNTLESFRENLDNDFDYNNKYRKRMEIY 180
Db 121 VNSDYNVNFYDSNAFLINTLTKTIVESAINDTTLESFRENLDNDFDYNNKYRKRMEIY 180
QY 181 DKQGHFDYKSOJTEENPEFIIDNIKTLYLSNEYSKDLEALNKYTEESLNKITANNNDI 240
Db 181 DKQKNFINYKAQREENPELLIDIVKTYLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240
QY 241 RNLEKPADDEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVDILPGIQDLFKS 300
Db 241 RNFEFKNGESFNLYEQELVERWNLAAASDILRISALKEIGMYLDVMDLPGIQDLFPES 300
QY 301 INKPDSTINTSWEMIKLEAIMKYKEYIPGYTSKNPMDLDEEVQSFESALSCKSKSEIF 360
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSEHFDMLDEEVQSFESVLSKSKSEIF 360
QY 361 LPLDDIKVSPLEVKIAPANNSVINOALISLSDKSCVSLVINOIKNRYKILNDNLNPSNE 420
Db 361 SSLGDMSEASPLEVKIAPFNSKGIINOGLISVSKDSYCSNLIVQKIEINRYKILNNSLPALSE 420
QY 421 GTDFTNTTKIFSDKLASINEDNMFMIKITNYLKVGFAPVPRSTINLSGPGVTGAYQD 480
Db 421 DNDFTNTTNTFIDSI MAEANADNGRFWMELGLRVGFFPDVKITINLSGPEAYAAQD 480
QY 481 LLMPKDNSTNHLLEPELNFEPFKTKISQTEQITSLMSFNQARAKSQEYKKGYPF 540
Db 481 LLMPKEGSMNTHLEADLRNFESIKNTNISQSTQEMASLWFSFDDARAKAQEYKRNYPF 540
QY 541 GALGEDDNLDPQNTVLDKDYVSKILSSMTKRNEYIHYIVOLQGDKI SYEASCNLFSK 600
Db 541 GSLGEDDNLDFSQNIIVDKKEYLLEKISLARSSEGYHYIVOLQGDKISYEAACNLPAK 600
QY 601 DPYSILYQKNIEGSETAYYYYVADAETKEIDKYRIPYQISNKRNIKLTFTIGHGSEFNT 660
Db 601 TPYDSVLFQKNIEDSEIAYYVPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNT 660
QY 661 DTFANLDVDSLSSEIETILNLAKADISPKYTEINLLGCNMFYSYIAYEETPGKLLLIK 720
Db 661 DIFAGFDVDSLSTEIAAIDLAKEDISPKYTEINLLGCNMFYSYINVEETPGKLLLVK 720
QY 721 DRVSELMPSISODSITVSANOYEVNRINEEGREILDHSGKWKINKEESI IKDISKEYISF 780
Db 721 DKISELMPSISQDSIIVSANQYEVNRINEEGREILDHSGEWINKKEESI IKDISKEYISF 780
QY 781 NPKENKIIVKSKYLHELSTLLOEIRNNANSDDILEKKVMLTECEINVASNIDRQIVGR 840
Db 781 NPKENKIIVKSKNLPSTLLQEIIRNNNSSDILEEKVMLTECEINVISNIDTQIVEER 840
QY 841 IEEAKNLTSDSINYIKNEFKLIESISLYDLKHQNGLLDSDHFSFEDISKTENGFRFRF 900
Db 841 IEEAKNLTSDSINYIKNEFKLIESISDALCDLKQONELEDSDHFSFEDISSETDEGFSIRF 900
QY 901 INKETGNSIFJTEKEIFESEYATHISKEISNIDTIFDNVNGKLVKYNLAAHEVNTLN 960
Db 901 INKETGESIFVETEKTI FSEYANHITESIKIGTIFDTVNGKLVKYNLDTTHEVNTLN 960
QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAOQLFSTGLNTITDASKVVVELYSTALDETID 1020

Db 961 RAFFIQSLIEYNSKESLNSLVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDETID 1020
Qy 1021 LPTLSEGLPIITATIDGVSIGAAIKELSETNDPLLRQIEAKIGIMAVNLTAARSTAVT 1080
Db 1021 LPTLSEGLPIITATIDGVSIGAAIKELSETSDPLLRQIEAKIGIMAVNLTAARSTAVT 1080
Qy 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELIOLQKATKVIDYFKHISLAETEGAFITLD 1140
Db 1081 SSGIASGFSILLVPLAGISAGIPSLVNNELVLRQATKVVDPFKHISLVETEGVFTLLD 1140
Qy 1141 DKIMPQDVLVSEIDFNNSITLQKCEIWRAGSGHTLTDIDHFFSSPSITVRKPNL 1200
Db 1141 DKIMPQDVLVSEIDFNNSITLQKCEIWRAGSGHTVTDIDHFFSAPSITVRPHL 1200
Qy 1201 SYDVLNITKKEKIDFSKOLVLPNAPNRPVFGEMGTGCFSLNDNDGTLKLDRIHDHVEG 1260
Db 1201 SYDVEVQKEELDSKOLVLPNAPNRPVFAVETGTPCLRSLENDGTLKLDRIHDHVEG 1260
Qy 1261 QPYWRYFAFIADALITKLPYEDTNVRINLGNTRSFIVPITTEQIRKNLSYFYSGG 1320
Db 1261 BPYWRYFAFIADALITKLPYEDTNVRINLGNTRSFIVPITTEYIREKLSYFYSGG 1320
Qy 1321 GYSLSLSPYNNIDLVNDTWIDVNVVKNVTIESDEIQKGLNLSKLNIEDN 1380
Db 1321 GTYALSLSQYNNGINELSESVDWIIDVNVVRDVTIESDKIKGDLIEGILSTLSIBEN 1380
Qy 1381 KIILNHTINFYGDINERFSLFSTILEDINIIIEIDLVSYSKILLSGCMKLIENS 1440
Db 1381 KIILNSHEINFGEVNGSGVSLFSLFLEGINAIIEVDLLSKYSKILLSGELKILMLNS 1440
Qy 1441 SDIOQKIDHIGFNGHQKIFPYSYIDNETKNGFDIDYSKKEGLTAEFNSIESIIRNIYMP 1500
Db 1441 NHIQKIDYIGNSELOKNIPIFVDSSEKENGPFNGTKEGLFVSELPDVVLISKVYMD 1500
Qy 1501 DNNLPYVSSKDLKDIRINKGDKVLLIGNYPKDMKVSLSFTIEDTNTIKLVGYLDEN 1560
Db 1501 DSKPSFGYVSSNNLKDKVITKDNVILTGYYLKDDIKISLSLTLODEKTIKLSVHLDES 1560
Qy 1561 GVAQILKFNNAKSALNTSILNMFLESINIKIFVNLDPNIEFTLDTNFIISGNSIG 1620
Db 1561 GVAELKFNMR-KGNTNTSDLSLMSLESNNIKSIFVNLQSNIKFILDNFIISGTSIG 1619
Qy 1621 QPELICDKDKNTQPYFINKETSYLVGNRQNLVPEPSYHLDOSGNSISSTVINFSOK 1680
Db 1620 QPEFICDENDNQPYFINKETLETNTYLVGNRQNLVPEPNYDLDSDGDISSTVINFSOK 1679
Qy 1681 YLYGIDRYNVKVIAPNLYT 1700
Db 1680 YLYGIDSCVNVKVISPNLYT 1699

RESULT 4
US-08-957-310-10
; Sequence 10, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-957-310-10

Query Match 77.6%; Score 6733.5; DB 4; Length 2366;
Best Local Similarity 76.4%; Pred. No. 0;
Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;

Qy 1 MNLVNAQLQKQVYVYKRIQEDYVAILNALAEYHNMSSESVKYLKLDKINNLTADYL 60
Db 1 MSLVNRKQLEKXNVYRFTQEDYVAILDALAEYHNMSSENTVVEKYLKLDKINSLTDIYI 60

Qy 61 NTKKSGRKNALKKPKVEYLTWVELEKNNSLTPVEKNLHFVIGWQINDTAINYNQWKD 120
Db 61 DTKKSGRKNALKKPKVEYLTWVELEKNNSLTPVEKNLHFVIGWQINDTAINYNQWKD 120

Qy 121 VNSDYTVKVFVDSNAFLINTLKKTIVESATNTNLTSPRENLDNDFDYNKPKRMEIYY 180
Db 121 VNSDYTVKVFVDSNAFLINTLKKTIVESATNTNLTSPRENLDNDFDYNKPKRMEIYY 180

Qy 181 DKQHFIDYKSOIEENPEFIIDNIITKLYSNEYSKOLEALNKYIESSLNKITANNNGDI 240
Db 181 DKQNFINYKQREENPELIIDDIKTYLSNEYSKIEDELNTYIESSLNKITQNSGNDV 240

Qy 241 RNLEKFADEDLVRLYNQELVERWNLAAASDLIRISMLKEDGGVYLDVILPGIQDLPKS 300
Db 241 RNLEKFADEDLVRLYNQELVERWNLAAASDLIRISMLKEDGGVYLDVILPGIQDLPKS 300

Qy 301 INKPSITNTSWEMIKLEAIKMYKEYIPGYTSKNFDMLEDEEVQSFESALSSKSDKSEIF 360
Db 301 IEKSSVTVDVFWMTKLEAIKMYKEYIPGYTSKNFDMLEDEEVQSFESALSSKSDKSEIF 360

Qy 361 LPDLDIKVSPLEVKIAFANNVINOALISLKDYSCLVINOIKNRYKILNDNLNPSINE 420
Db 361 SSGIDMEASPLEVKIAFNKGIINGQLISVKDYSCLVINOIKNRYKILNDNLNPSINE 420

Qy 421 GTDFNTTKKIFSDKLASISNEDNMFMKIITNYLVKGFADPVRSSTINISGPGVYTGAYQD 480
Db 421 DNDFTNTTNTFIDSIMAEANADNGRFMMELGKLVFGFFPDVYKTTINLSGPEAYAAAYQD 480

Qy 481 LLMFKDNSTNHLLEPELNRNEFPKTKISQLEQITSLMSFNQARAKSOEEYKGYFE 540
Db 481 LLMFKEGSMNHLLEPELNRNEFPKTKISQLEQITSLMSFNQARAKSOEEYKGYFE 540

Qy 541 GALGEDDNLDFQNTVLDKDYVSKKILLSGSMKTRKEYIHYIVLOQDKISYEASCNLFPSK 600
Db 541 GALGEDDNLDFQNTVLDKDYVSKKILLSGSMKTRKEYIHYIVLOQDKISYEASCNLFPSK 600

Db 541 GSLGEDNDLFSQNIIVVDKEYLLEKISSLARSSERYIHYIVLQOQDKISYEACNLFAK 600
 Qy 601 DVSSTLYOKNTEGSTAYVYVADAEIKEDIKYRIPYOISNKRNIKLTFFIGHGSEFNT 660
 Db 601 TPYDSVLFQKNIEDSIAIYNPGDGEIDIKYKIPSIIDRPKILTFIGHGDEFT 660
 Qy 661 DTFANLDVDSLSSEIETILNLAADIPKPIYIEINLLGCNMFYSIYAETYPGKLLKIK 720
 Db 661 DIFAGFDVDSLSSEIETILNLAADIPKPIYIEINLLGCNMFYSIYAETYPGKLLKIK 720
 Qy 721 DRVSELMPSISQDSITVSANQVEVRINERKREIIDLHSGKWNKESIIKDISKEYISF 780
 Db 721 DKISELMPSISQDSITVSANQVEVRINERKREIIDLHSGKWNKESIIKDISKEYISF 780
 Qy 781 NPEKNKIIVKSVLHSLTLLQEIIRNANSDDIDLEKKVMTCEINVASNDROIYGR 840
 Db 781 NPEKNKIIVKSVLHSLTLLQEIIRNANSDDIDLEKKVMTCEINVASNDROIYGR 840
 Qy 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDDSHFISFEDISKTEGFRIRF 900
 Db 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDDSHFISFEDISKTEGFRIRF 900
 Qy 901 INKETGNSIFETEKEIFSEYATHISKEISNITKOTIFDQVNGKLVKQVNLDAAEHVNTLN 960
 Db 901 INKETGNSIFETEKEIFSEYATHISKEISNITKOTIFDQVNGKLVKQVNLDAAEHVNTLN 960
 Qy 961 SAFFIOTSLIYNTKESLNSLVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDETID 1020
 Db 961 SAFFIOTSLIYNTKESLNSLVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDETID 1020
 Qy 1021 LPTLSEGLPIATIIDGVSGLAAIKELSETDPLLRQIEAKIGIMAVNLTAATAIT 1080
 Db 1021 LPTLSEGLPIATIIDGVSGLAAIKELSETDPLLRQIEAKIGIMAVNLTAATAIT 1080
 Qy 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETEGFTLLD 1140
 Db 1081 SSGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVIDYFKHISLAETEGFTLLD 1140
 Qy 1141 DKIMPODDLVLSEIDFNNSITLGKCEIWRAGSGGHTLTDIDHFFSPSITVRKPL 1200
 Db 1141 DKIMPODDLVLSEIDFNNSITLGKCEIWRAGSGGHTLTDIDHFFSPSITVRKPL 1200
 Qy 1201 SIYDVNLNKEKIDFSDKMLVLPNAPNRFVFGEMGTGFRSLDNDGTLKLRDRHYEG 1260
 Db 1201 SIYDVNLNKEKIDFSDKMLVLPNAPNRFVFGEMGTGFRSLDNDGTLKLRDRHYEG 1260
 Qy 1261 QFYWRVFAFIADALITKLPREDTNRINLGNTRSFIVPITTEQIKNLSYFYSGG 1320
 Db 1261 EFWYRFAFIADALITKLPREDTNRINLGNTRSFIVPITTEQIKNLSYFYSGG 1320
 Qy 1321 GSYLSLSYNNMIDNLVENDTWIDVNVVKNITIESEDEKQELIENILSKLNIEDN 1380
 Db 1321 GSYLSLSYNNMIDNLVENDTWIDVNVVKNITIESEDEKQELIENILSKLNIEDN 1380
 Qy 1381 KIILNHTINFYGDINESRFLSLEIDNIIIEIDLVSQYKILLSGCMKLIENS 1440
 Db 1381 KIILNHTINFYGDINESRFLSLEIDNIIIEIDLVSQYKILLSGCMKLIENS 1440
 Qy 1441 SDIQKIDHIGFNGHQKIFYSYIDNETKNGFYDYSKKEGLFTAFESNESIIRNIYMP 1500
 Db 1441 NHIQKIDHIGFNGHQKIFYSYIDNETKNGFYDYSKKEGLFTAFESNESIIRNIYMP 1500
 Qy 1501 DSNNLFIYSKDKOTRIINKGDKVLLIGNYFKDDMKVLSFTIEDNTIKLNGVVLDE 1560
 Db 1501 DSNNLFIYSKDKOTRIINKGDKVLLIGNYFKDDMKVLSFTIEDNTIKLNGVVLDE 1560
 Qy 1561 GVAQILKFMNNAKSALENSLNFLESINIKNIFYNLDPNIEFTLDTNFTISGNSGT 1620
 Db 1561 GVAQILKFMNNAKSALENSLNFLESINIKNIFYNLDPNIEFTLDTNFTISGNSGT 1620
 Qy 1621 QFELICDQKNOPIYFINKETSYTLVYGNRQNLIVPSPVHLDDSGNISSTVINFSQK 1680
 Db 1620 QFELICDQKNOPIYFINKETSYTLVYGNRQNLIVPSPVHLDDSGNISSTVINFSQK 1680

Qy 1681 YLYGIDRYVNVKVIAPNLYT 1700
 Db 1680 YLYGIDSCVNVKVIAPNLYT 1699
 RESULT 5
 US-10-011-366-10
 ; Sequence 10, Application US/10011366
 ; Patent No. 6573003
 ; GENERAL INFORMATION:
 ; APPLICANT: Williams, James A.
 ; Kink, John A.
 ; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
 ; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
 ; DISEASE
 ;
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Medlen & Carroll
 ; STREET: 220 Montgomery Street, Suite 2200
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/011.366
 ; FILING DATE: 16-NO. 6573003-2001
 ; CLASSIFICATION: <unknown>
 ;
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/957.310
 ; FILING DATE: 23-OCT-1997
 ; APPLICATION NUMBER: US 08/329.154
 ; FILING DATE: 24-OCT-1994
 ; APPLICATION NUMBER: US 08/161.907
 ; FILING DATE: 02-DEC-1993
 ; APPLICATION NUMBER: US 07/985.321
 ; FILING DATE: 04-DEC-1992
 ; APPLICATION NUMBER: US 07/429.791
 ; FILING DATE: 31-OCT-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ingolia, Diane E.
 ; REGISTRATION NUMBER: 40,027
 ; REFERENCE/DOCKET NUMBER: OPD-01121
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 705-8410
 ; TELEFAX: (415) 397-8338
 ;
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2366 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
 ;
 ; US-10-011-366-10
 ;
 ; Query Match 77.6%; Score 6733.5; DB 4; Length 2366;
 ; Best Local Similarity 76.4%; Pred. No. 0;
 ; Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;
 ;
 ; Qy 1 MNLVNAQLOKQVYKFIQDEYVAIILNALLEEYHNMSESSVVEKYLKDKOINLTNYL 60
 ; Db 1 MSLVNRKQLEKMANVFRTOQDEYVAIILDALEEYHNMSESSVVEKYLKDKOINLTNYL 60
 ; Qy 61 NTKYKSGRKNALKKFPEKYLTMVELELKNSTLPVEKNLHFHWIGGQINDTAINYNQWKD 120
 ; Db 61 DTKYKSGRKNALKKFPEKYLTMVELELKNSTLPVEKNLHFHWIGGQINDTAINYNQWKD 120
 ; Qy 121 VNSDYTVKRVYDSNAFLINTLTKTIVESATNTNTLESFRENLDNDFDYNKFKYRKRMEIY 180

Db 121 VNSDYNVNFYDSNAFLINTLTKTVVESAINDTLESFRENLDPRDYNNKFRKMEIY 180
Qy 181 DKQKHFIDYKSOIENPEFIIDNIITKTVLSNEYKOLBAUNKYIEESLINTKITTANGNDI 240
Db 181 DKQKHFIDYKSOIENPEFIIDNIITKTVLSNEYKOLBAUNKYIEESLINTKITTANGNDI 240
Qy 241 RNLEKFADEDLVRLNQEELVERWNLAAASDIIRISMLKEDGGVYLDVILPGIQDLPKS 300
Db 241 RNLEKFADEDLVRLNQEELVERWNLAAASDIIRISMLKEDGGVYLDVILPGIQDLPKS 300
Qy 301 INKPSITNTSWEMIKLEAIMKYIPIGYTQKFNDFMLDEEVQSFESALSKSKSEIF 360
Db 301 INKPSITNTSWEMIKLEAIMKYIPIGYTQKFNDFMLDEEVQSFESALSKSKSEIF 360
Qy 361 LPDLDIKVSPLEKVFAPANNVINQALISLKDYSQDIIVINOIKRKYILNDNLKPSINE 420
Db 361 LPDLDIKVSPLEKVFAPANNVINQALISLKDYSQDIIVINOIKRKYILNDNLKPSINE 420
Qy 421 GTDFNTMTKIFGDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYQD 480
Db 421 GTDFNTMTKIFGDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYQD 480
Qy 481 LLMFKDQNSTNHLBPELNFPPKTKISQLEQETISLWSFNQARAKSQPEYKGYE 540
Db 481 LLMFKDQNSTNHLBPELNFPPKTKISQLEQETISLWSFNQARAKSQPEYKGYE 540
Qy 541 GALGDDNDLFAONTVLDKDYVSKILSSMKTRNKEYIHYIYVLOQDKISYEASCNLPSK 600
Db 541 GALGDDNDLFAONTVLDKDYVSKILSSMKTRNKEYIHYIYVLOQDKISYEASCNLPSK 600
Qy 601 DPYSILYOKNIEGSETAYYYVADAIEIKIDKIRIPIQISNKRNIKLTFFIGHKSEFNT 660
Db 601 DPYSILYOKNIEGSETAYYYVADAIEIKIDKIRIPIQISNKRNIKLTFFIGHKSEFNT 660
Qy 661 DTFANLVDLSSEIETILNLAADISPKYIIEINLGCNMFYSIYAEITYPGKLLKIK 720
Db 661 DTFANLVDLSSEIETILNLAADISPKYIIEINLGCNMFYSIYAEITYPGKLLKIK 720
Qy 721 DRVSELMPSISODSITVSANOYEVNINEBKEIIDLHSGKWKNEESIIKDISKEYTSF 780
Db 721 DRVSELMPSISODSITVSANOYEVNINEBKEIIDLHSGKWKNEESIIKDISKEYTSF 780
Qy 781 NPKENKIIVKSKYHELSTLLOEIRNANSIDILEKKVMTCEINVASNIDROIVEER 840
Db 781 NPKENKIIVKSKYHELSTLLOEIRNANSIDILEKKVMTCEINVASNIDROIVEER 840
Qy 841 IBEAKNLTSDSINYIKNEFKLIESDSLYDLKHQGLDSDHPIFSDISKTENGFRIRP 900
Db 841 IBEAKNLTSDSINYIKNEFKLIESDSLYDLKHQGLDSDHPIFSDISKTENGFRIRP 900
Qy 901 INKETSNSIFITEKEIFSEYATHISKEISNTKDTIFDNVNGKLVKKVNLDAHEVNTLN 960
Db 901 INKETSNSIFITEKEIFSEYATHISKEISNTKDTIFDNVNGKLVKKVNLDAHEVNTLN 960
Qy 961 SAFFIQLSEYNTTESLNSLVAMKVQVYAOVLFSTGLNTIITDASKVVELVSTALDETID 1020
Db 961 SAFFIQLSEYNTTESLNSLVAMKVQVYAOVLFSTGLNTIITDASKVVELVSTALDETID 1020
Qy 1021 LPLTISEGLPIIATIIDGVSGLAAIKELSETNDPLLRQIEAKIGIMAVNLTAATAIYT 1080
Db 1021 LPLTISEGLPIIATIIDGVSGLAAIKELSETNDPLLRQIEAKIGIMAVNLTAATAIYT 1080
Qy 1081 SALGTASGSIILVPLAGISAGIPSLVNNELLLODKATKVIDYFKHISLAETEGFTLID 1140
Db 1081 SALGTASGSIILVPLAGISAGIPSLVNNELLLODKATKVIDYFKHISLAETEGFTLID 1140
Qy 1141 DKIMPDQDLVISEIDFNNSITLKGCEIWRAGSGHGLTDIDHFFSPSPITRKPLV 1200
Db 1141 DKIMPDQDLVISEIDFNNSITLKGCEIWRAGSGHGLTDIDHFFSPSPITRKPLV 1200
Qy 1201 STYDVLNKKKIDFSKOLMVLNPNAPRVFGEMGWTGFRSLDNDGTKLLDIRIDHYEG 1260
Db 1201 STYDVLNKKKIDFSKOLMVLNPNAPRVFGEMGWTGFRSLDNDGTKLLDIRIDHYEG 1260

Db 1201 STYDVLVEQBELDLSKDLAMVLPNAPRVFAWETGWTGFRSLDNDGTKLLDIRIDHYEG 1260
Qy 1261 QFYWRYPAFIADALITTLKPRYEDTNVINLDGNTRSFIVPITTEQIRKKNLSYFYSGG 1320
Db 1261 QFYWRYPAFIADALITTLKPRYEDTNVINLDGNTRSFIVPITTEQIRKKNLSYFYSGG 1320
Qy 1321 GSYLSLSPPYNNIDNLVENDTWVIDVNVVKNITTESDEIQKGLIENILSKLNIEDN 1380
Db 1321 GSYLSLSPPYNNIDNLVENDTWVIDVNVVKNITTESDEIQKGLIENILSKLNIEDN 1380
Qy 1381 KIILNNHTINPYGDINESNRFISLTFSTLEIDNIIIEIDLVSYSKILLSGNCMKLIENS 1440
Db 1381 KIILNNHTINPYGDINESNRFISLTFSTLEIDNIIIEIDLVSYSKILLSGNCMKLIENS 1440
Qy 1441 SDIOQKIDHIGFNGSHQKVIYFVSYIDNETKNGFIDYKKEGLTFAEFSNESIIRNIYMP 1500
Db 1441 SDIOQKIDHIGFNGSHQKVIYFVSYIDNETKNGFIDYKKEGLTFAEFSNESIIRNIYMP 1500
Qy 1501 DSNNLFIYSSKDLKDIRIINKGDVKKLIGNYFKODMKVSLSFETIEDTNTIKLNGVYLDEN 1560
Db 1501 DSNNLFIYSSKDLKDIRIINKGDVKKLIGNYFKODMKVSLSFETIEDTNTIKLNGVYLDEN 1560
Qy 1561 GVAQILKPMNAKSMALNTSNLSMNPLESINIKIPYNNLDPNIEFILDNTNFIISGNSIG 1620
Db 1561 GVAQILKPMNAKSMALNTSNLSMNPLESINIKIPYNNLDPNIEFILDNTNFIISGNSIG 1620
Qy 1621 QFELICDCKDKNIQPYFINFKIKETSYTLVGNRQNLVPEPSVHLDDSGNISSTVINFSOK 1680
Db 1621 QFELICDCKDKNIQPYFINFKIKETSYTLVGNRQNLVPEPSVHLDDSGNISSTVINFSOK 1680
Qy 1681 YLYGIDRYNVKVIAPNLYT 1700
Db 1681 YLYGIDRYNVKVIAPNLYT 1700
Qy 1690 YLYGIDSCVKNVWISPIYT 1699
Db 1690 YLYGIDSCVKNVWISPIYT 1699
RESULT 6
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994

US-08-405-496A-6
; Sequence 6, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-405-496A-6

Query Match 51.4%; Score 4458; DB 2; Length 2710;
Best Local Similarity 50.7%; Pred. No. 1.5e-261;
Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;
QY 1 MNLVNAQOLQKVVYKFRQDEYVAILNALREYNMSSVVEKYKLKDLNLTNYL 60
DB 1 MSLISKEELIKUAY-SIRPRENEYKLTILNDEYNKLTNNENKYLQKUNESIDVPM 59
QY 61 NTYKSGRNKALKKPKXYLTMEVLKNNSLTPVEKNLHFIWGGQINDTAINYNQWKD 120
DB 60 NKYKTSRRNALSNLKDKILKEVILKNSNTSPVEKNLHFVWIGEVSDIALEYIKQWAD 119
QY 121 VNSDYTVKVPYDSNAPLINTLTKTIVESATNTNLTSEFRNLNDPDPYKFRKRMETIY 180
DB 120 INAEYNIKLYWSEAFVNLTKKAIIVESSTTEALQLEBEIQNPQPDNNKFKYKGRMEFY 179
QY 181 DKQKHFIDYKQISENPEFIIDNIKTVLSNEYSKDLKALNKYIEESLKNKITANNNDI 240
DB 180 DRQKRPINYYKQINRPVPTIDDIKSHLVSEYNRDETVELSEYRTNSLRKINSNHGDI 239
QY 241 RNLEKFAEDDLVRLNQELVERWNLAASDIIRISMLKEDGGVYLDVILPGIQPDLFKS 300

DB 240 RANSLFTEQELLNTIYSQELLNRGNLAAASDIIVRLALKNFGGVYLDVMDLPGIHSDFLPT 299
QY 301 INKPDSTINTSWEMIKLEAIWKYKEYIPCYTSKNPDMLEDEEVQSFESALSKSKSSEIF 360
DB 300 ISRPSICGLDRWEMIKLEAIWKYKYNNTYSENFDKLDQOKNFKLIIESKSEKSEIF 359
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLKDSCYSDIIVINOIKRNYKILNDNLNPSINE 420
DB 360 SKLENLNVSDLEIKIAPALGSVINOALISKQSYLTNLVIBQVKRYQFLNQHLPAYES 419
QY 421 GTDFTNTMKIPSDKLASISNEDNMFMKIITNYLKVGFAPDVRSTINISGCVYTGAYQD 480
DB 420 DNNFTDTTKIHFDSLFSNATAENSMTKIAPYLOVGFMPPEARSTISLUSGAYASAYD 479
QY 481 LLMPKDNSTNIHLEPELNRFPFKTKISQLTQOEITSLWSFNQARAKSQPEYKKGYPE 540
DB 480 FINQENTIEKTLKASDLIEKFPENNLSQLTQEINSLWSPDQASAKYQEKYVRYDTG 539
QY 541 GALGEDDNLDPQONTVLDKQY-VSKKILSS--MTRNKKEYIHYIVOLGDKIISYASCNL 597
DB 540 GSLSEDNGVDFNKNTALDKNYLLNNKIPSNVVEAGSKNYVHYIIQLQDDISYEATCNL 599
QY 598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKEDKVRIPYQISNKNIKLTFIGHCK 655
DB 600 FSKNPKNSIIQRNN--NESAKSYPLSDGSEISLELNKYRIPERLKNKEKVKVTFIGHGK 657
QY 656 SEFNTDTEANLDVDSLSEIETILNLAKADISPKYIBINLGCNMFSYSIYAEETYPGKL 715
DB 658 DEFNTSEFARUSVDSLSEISSEIFDTIKDLSPRNVEVNLGCNMFSYDFNVEETYPGKL 717
QY 716 LLKIKDRVSELMPGISQDSITVSANQYEVRIINEEGKREILDHSGKWIKEESIIDKISSK 775
DB 718 LLSIMDKITSTLPDVNKNSTIGANQYEVRIINSEGRKELLAHSGKWIKEEAIMSDLSK 777
QY 776 EYISFNPKENKIIKVKYKYLHELSTLQLOIRNANSOIDLKKNVMTCEINVAENIRQ 835
DB 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLILDASVSPDTKTLNKLKNISSIGDY 837
QY 836 IVEGRIEAKNLTSDSINVYKNEPKLIESISDLSYDLKHQGLDDSHFTSPEDIKSTENG 895
DB 838 IYYEKLEPVKNIIHNSIDDLDEFNLEENVSDELYELKLNLDKEYLISFEDISEKNST 897
QY 896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNIKDTIPDNVNGKLVKKNVLDAAHE 955
DB 898 YSVRPINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGNLDNIQLDHTSQ 957
QY 956 VNTLNSAFPIQSILBYNTTKESLNSLVAMKVQVYLAQLPSTGLNTITDASKVVELVSTAL 1015
DB 958 VNTLNAAPFIQSLIDYSSNKDVLNLTSTSVKQVLAQLFSTGLNTIYDSIQVNLISNAV 1017
QY 1016 DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPDLRQETAEKIGIMAVNLTAAS 1075
DB 1018 NDTINVLTPIITTEGPIVSTIILDGINIGAAIKELDEHDPDLKKELEAKGVGLAINNSLSI 1077
QY 1076 TAITVSALGIASGFSILLVPLAGISAGIPSLVNNELIQLDKATKVIDYFKHISLAETEGA 1135
DB 1078 AATVASIVGIGAEVTFILPPIAGISAGIPSLVNNELILHDKATSVVNFHLSSEKCYCP 1137
QY 1136 FTLLDDKIIMPDQDLVLSEIDFNNNSITLKGCEIWRAGGGSGHTLTDDIDHFFSPSPITY 1195
DB 1138 LKTEDDKILVPIDDLVISEIDFNNNSIKLGTGNILAMEGGSGHTVTGNIDHFFSPSPISS 1197
QY 1196 RKPWLSIYDVLNLIKKEKIDFSDKOLMVLNAPNRFVGYEMGTTPGFRSLDNDGCTKLLDR 1255
DB 1198 HIPSLSIYSAIGIETENLDFSKKIMMLPNAPSRVFWNETGAVPGLSRLENDGCTRLDLSIR 1257
QY 1256 DHYEQFYWRYPFAITADALITKLKRYEDTNNRINLDGNTRSFIVPVITTEQIRKNLSYS 1315
DB 1258 DLYPKFWRFYAFYF-DYAITTLKVPYEDTNIKIKLDKDRNFIMFTTITNIRNKLVSYS 1316
QY 1316 FVSGGGSYSLSPYNNMIDNLNVENTWIDVDNVKNITTESDEIQGELIENILSKL 1375
DB 1317 PDGAGGTYSLLSSYPISNTINLSKDDLWIFNIDNEVRBISIENTGKIKGKLIKDVLSKI 1376

QY 1376 NIEDNKILNHTINPYGDINESRFLSTFSTILEDINIIIEIDLVSXYKILLSNCWK 1435
Db 1377 DINKNLIIGNQDTISGDDINKDRIFUTCEDDKISLIIIEINLVAKSYLLSGDKNY 1436
QY 1436 LIENSSDIOQKHIDHGFNGEHOKYIFYSYID-NETKYNFIDYSKKEGLFTAEPNSIESII 1494
Db 1437 LISNLSNTEKINTGLD--SKNIAVYVDESNNKYFCAL-----SKTSQKSII 1483
QY 1495 RNIYMPDSNNL-----FIYSSKDL--KDIRINKGVDKLLIGNYFKD---DMKVSLS 1541
Db 1484 H--YKDSKNILEFYNDSTLEFNSKDFIAEDINVFMKDDINTITGKYVVDNNTDKSIDFS 1541
QY 1542 FRIEDTNTIKLGVYLDENGVAQILKFMNNAKSALNTSNLAFLESINIKNIFNNLDP 1601
Db 1542 ISLVSKNQVKNGLYLVSYSLDFVKNDSOHHNTSNFNLFDNISFWLKGFPE--- 1598
QY 1602 NIEFILDNTFIISGNSIQGFELICDKOKNIQPIFYFINFKIKETSVLYVGNRQNLIVBPS 1661
Db 1599 NINFVIDKYFTLVGKTNLGVVEFICDNNKNIIDYFGEWKTSKSTIFSGNGRNVVVEPI 1658
QY 1662 YHLDGSGNSTVINSQKYLGVIDRYVKNVIAPNLYT 1700
Db 1659 YN-PDTGEDISTSLDPSYBPLYGIDRYINKVLIAPDLYT 1696

RESULT 8

US-08-915-136-6
Sequence 6, Application US/08915136
Patent No. 6290960
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-6

Query Match 51.4%; Score 4458; DB 3; Length 2710;
Best Local Similarity 50.7%; Pred. No. 1.5e-261;
Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;

QY 1 MNLVNAKQLOKMYVVFRIQDEYVAIILNAAEEYHNHNSSESVVEKYLKLDINNLTNLYL 60
Db 1 MSLISKEELIKLAY-SIRPRENEYKTLTNDVYKNLTNNNNENKYLQKKLINESIDVFM 59
QY 61 NTYKSGRNKALKKPKFKEYLTMEVLELKNLSLTPVEKNLHFIIWIGQINDTAINYNQWKD 120
Db 60 NKYTSSNRALSNLKDLKEVILIKNSNTPVEKNLHFVWIGGEVSDIALEYIKQWAD 119
QY 121 VNSDYTVKVVDSNAFLINTLTKTIVESATNNTLESFRENLDNPDFFYKRYKMEIY 180
Db 120 INAEYNIKLYDSEAFVLTAKAIVESSTTEALQLEEEIQNPQFDMNKYFKRMEFY 179
QY 181 DKQKHFDYKSOJTEENPEFIIDNIIKTYLSNEYSKDLKALNKYIEESLNKITTANGNDI 240
Db 180 DRQKRFINYKSOQNKPTVPTIDDIILKSHLVSEYNRDETIVLESYRTNSLRKINSNHGDI 239
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGCVYLDVILPGIQDPLFS 300
Db 240 RANSLFTEQELLNYSQELLNARGLNLAASDIVRLLAKNFGVYLDVMDLPGIHSDLFT 299
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKPMDEEVOVSFEALSCKSKSEIF 360
Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYINNYTSENFDKLDQOLKONFKLIIESKSEIF 359
QY 361 LPDDIKVSPLEVKIAPANNVINOALISLSDVCSDLVINQIKRYKILNDNLNPSINE 420
Db 360 SKLENLVSDLEIKIAFGVINOALISKOSYTLNLVIEQVKRYQLNQHLPALIES 419
QY 421 GTDFNTTMKIFSDKLASINEDNMFMIKITNLYKFGFAPDVRSTINISGPGVYTGAYQD 480
Db 420 DNNFTDTTKIFHDSLFNSATAENSMTLKIAPYLOVGFMPPEARSTISLGGAYASAYD 479
QY 481 LLMFKNSTNTHLLEPELNRNFEFPKTKISQTEQBITSLWFSNQARAKSQEYKKGYPF 540
Db 480 FINLQNTIEKTLKASDLIEFKFPENNLSQLTEQBEINSLWSFDQASAKYQEKYVRDVTG 539
QY 541 GALGEDNDLFAQNTIVLDKDY-VSKKILSS--MKTRNKEYIHYIVOLQGDKISYASCNL 597
Db 540 GSLSEDNGVDNFKNTALDKNLLNNKIPSNVVEAGSKNYHYIILQGGDISYBATCNL 599
QY 598 FSKDPYSSILYQKNIEGSETAYVYVAD--AEIKEDKYRIPYQISNRKNLKTIFIGHK 655
Db 600 FSKPKNSIIIIQRNM--NESAKSVFLSDGDSILELNKRYRPERLNKKEKVKVTFIGHK 657
QY 656 SEFNFTDTFANLDVDSLSSEIETILNLAkadISPKYIEINLGCNMFYSYIVAEETYPGL 715
Db 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLDISPKNVEVNLGCMNMFSDVFNVEETYPGL 717
QY 716 LLKIKDRVSELMPSISQDSITVSANOYEVNRNEEKREILDHSGKWNKESIIIDKSSK 775
Db 718 LLSIMDKITSTLPDVNKNISITIGANQYEVNRSEGRKELLASHGKWNKEAIMDLSK 777
QY 776 EYISFNPKNKIIIVKSKYLHELSTLLOQIRNANSDDIDLEKVMMLTCEINVASNIDRQ 835

Db 360 SKLENLVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHLPABIS 419
 Qy 421 GTDPNTMTKIFSKLASISNEDNMFMWIKITWYLVKGFAPDVRSTINLSGRGVYTCAYOD 480
 Db 420 DNNFTDTTKIFHDSLFNSATAENSMFLTIAPYLVQGFMEPEARSTISLSGPGAYASAYD 479
 Qy 481 LLMFKDNSTNIHLLPELANFEPKTKISQLTQEQEITSLWSFNQARAKQFBEYKKGYPE 540
 Db 480 FINLQENTYKTLKASDLLEFPENNLSQLTQEQEINLSWSDQASAKYQFKEYYRDYTG 539
 Qy 541 GALGEDNDLFAQNTVLDKDY-VSKKILSS--MKTRNKYIHYIYVQLQDKISYEASCNL 597
 Db 540 GSLSEDNGVDENKNTALDKNYLLANNKIPSNVVEEAGSKNYVHYIIQLQGDDEIYEATCNL 599
 Qy 598 FSKDPYSSILYQKNVIGSETAYYYVAD--AEIKETDKYRIPYQISNKNKNIKLTPIGHCK 655
 Db 600 FSKNPKNSIIQIRM--NESAKSYFLSDDBGESLELNLKRYRIPERLKNKEKVKVYTFIGHCK 657
 Qy 656 SEBNTDTFANLVDLSLSSETIETLNLAKADISPKYTEINLLGNMFSYSIYAEETYPGKL 715
 Db 658 DEFTSEFARLSVDSLNSISSFLDTIKLDISPANVEVNLGNMFSYDFNVEETYPGKL 717
 Qy 716 LKIKDRVSELMPISIQDSITSANQYEVRIINEEGREILDHSGKWINKEESIIDKISSK 775
 Db 718 LLSIMDKITSTLPDVNKNSTIGCANQYEVRIINSEGRKELLAHSGKWINKEEAIMSDLSK 777
 Qy 776 EYISFPNKENKIIVKSKYLHELSTLQEQEIRNNANSSDIDLEKKVMLTECEINVASNIDRQ 835
 Db 778 EYIFFDSIDNKLKAKSNIPGLASISEDIKTLDDASVSPDTKFIPLNNLKINIESIGDY 837
 Qy 836 IVEGRIEAKNLTSDINTYKNEFKLTIESIDSLYDLKHQGLDDSHFISFEDISKTEG 895
 Db 838 IYXKLEPVKNIIHNSIDLDIDFENLENSVDLEYELKLNLDKYLISFEDISKNST 897
 Qy 896 FRIRFINKETGNSIFIETEKEIPSEYATHISKEISNIKOTIFDNVNGKLVKVNLDAAHE 955
 Db 898 YSVRFINKNGESVYVETEKEIPSKYSEHITKEISTIKNSIITDVNGLNLDNIQDHTSQ 957
 Qy 956 VNTLNSAFTQSLIENYTKESLSNLSVAMKVQVYAQLSTGLNTITDASKVELYSTAL 1015
 Db 958 VNTLNAAFFQSLIDYSSNKNVDLNDLSTSVKQVLAQLFTGLNTIYDLSQVLNLSNAV 1017
 Qy 1016 DETIDLPLTUSEGLPIATIIDCVSLGAATKELSETNDPLLRQIEBAKIGIMAVNLTAAS 1075
 Db 1018 NDIINVLPITTEGIPVISTIDGINLOAIKELDEHDPDLKELBAKGVIAINMSLSI 1077
 Qy 1076 TAVTSALGIASGFSILLVPLAGISAGIPSLVNNELILOQKATKVIDYFKHISLAETEGA 1135
 Db 1078 AATVASIVGIGAEVTFILPIAGISAGIPSLVNNELILHDKATSVVNYFNHLSSEKKYGP 1137
 Qy 1136 FTLLDDKIIMPQDDVLSEIDFNNSITILGKEIWRAGSGHLLTDDIDHFPSSPSITY 1195
 Db 1138 LKTEDDKILVPIDDLVISEIDFNNSIKLGTGNILAMEGSGHVTGNDHFPSSPSISS 1197
 Qy 1196 RKPWLSYDVLNIIKKEIDPSKDLMLVLPNAPRVFVEMGWTGPFSLONGDKLDRJR 1255
 Db 1198 HIPSLSIYSAIGETENLDSKIMMLPNAPRVFWETGAVPGLSLENDGTGLLDSIR 1257
 Qy 1256 DHYEGOFYRYPAFIADALITKLKPRYEDTNVRINLDGNTSRFVITTEQIRKNLSYS 1315
 Db 1258 DLYPGKFWRFYAFF-DYAITTLKPVYEDTNIKIKLDDKOTRNFIMPTITTNNEIRNKL 1316
 Qy 1316 FYCGSGSYLSLSPYNNMIDLNVENDTWIDVNVVKNITTESDIOKGEIENITLSKL 1375
 Db 1317 FOGAGTYSLLSSYPSTINISKODLWIFNIDNEVREISENGITKKGKLIKOVLSKI 1376
 Qy 1376 NIEDNKIILNNHTINFGDINESNRFISLTFEIDNIIIEIDLVSYSKYLISGNCMK 1435
 Db 1377 DINKNLIIGNQIDISGDDINDKDRYFLTCLEDDKISLIIENLVAKSYLLSLSGDKNY 1436
 Qy 1436 LIENSDDIOQIDHIFGNGHQKIYFYSYID-NETKYNQFIDYSKKEGUFTAEFNSNESII 1494
 Db 1437 LISNLSNTIEKINTGLD---SKNIAYNYTDESNNKYFGAI-----SKTSQSKII 1483

Qy 1495 RNIVMPDSNNL-----FTYSSKDL--KDIRIINKGDKVLKLLIGNYFKD---DMKVSL 1541
 Db 1484 H--YKDSKNILEFVNDSTLEFNSKDFTAEDINVMFKDDINTYTKYVDNNTDKSIDFS 1541
 Qy 1542 FTIEDTNTIKLVGVLDENGVAQLKFMNNAKSALNTSNLMNFLESINIKNIFYNLDP 1601
 Db 1542 ISLVSKNQVKNGLYNESVSYLDFVKNSDGHHTSNFMNLFLDNISFWKLFGFE--- 1598
 Qy 1602 NIEFLDTNFIISGNSIGQFELICDKDKNQTPYFINFKIKETSITLYVGNRQNLIVEPS 1661
 Db 1599 NINFVIDKYFTLVGKTNLGYVEFICDNNKNIDIFYGEWKTSKSTISFGNGRNVVEPI 1658
 Qy 1662 YHLDGSGNISSTVNFNSOKYLYGIDRYVYKVIAPNLT 1700
 Db 1659 YN-PDTGBDISTLDSFSEPLYGIDRYINKVLIAPDLTY 1696

RESULT 10

US-10-011-366-6
 ; Sequence 6, Application US/10011366
 ; Patent No. 6573003
 ; GENERAL INFORMATION:
 ; APPLICANT: Williams, James A.
 ; Kink, John A.
 ; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
 ; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
 ; DISEASE
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Medlen & Carroll
 ; STREET: 220 Montgomery Street, Suite 2200
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/011,366
 ; FILING DATE: 16-NO. 6573003-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/957,310
 ; FILING DATE: 23-OCT-1997
 ; APPLICATION NUMBER: US 08/329,154
 ; FILING DATE: 24-OCT-1994
 ; APPLICATION NUMBER: US 08/161,907
 ; FILING DATE: 02-DEC-1993
 ; APPLICATION NUMBER: US 07/985,321
 ; FILING DATE: 04-DEC-1992
 ; APPLICATION NUMBER: US 07/429,791
 ; FILING DATE: 31-OCT-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ingolia, Diane E.
 ; REGISTRATION NUMBER: 40,027
 ; REFERENCE/DOCKET NUMBER: OPHD-01121
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 705-8410
 ; TELEFAX: (415) 397-8338
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2710 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-10-011-366-6
 Query Match. 51.4%; Score 4458; DB 4; Length 2710;

Qy	1437	I--ENSSDIOQKIDHIFNGEHBQKQIFVYSYIDNTKYNPFIDYSK----	KEGLFTAEPFN	1490
Db	1803	TFKEGINSLSSEALEHMN-----IDAIMSVIGLVQYARMIKNDNDISAIDHAG		1849
Qy	1491	E-SIIRNIYMPDSNNLFIYSSKDLKDIRINKGDVK-LLIGNYFKDDMKVLSFTIEDTN		1548
Db	1850	AVSDIKNIV-----DKFLGGTLTLTNRVYVPGVSGASLEGFISSGLEVCAS-----		1897
Qy	1549	TIKLVG---YLDENGVAQILKFMNNAKSALNTNSLMNFLESINIKNIFVNNLD-----		1600
Db	1898	--RMGGTAGRYL--SNVAKVIK-----PLLDI--GINWSLYDSSLNHAKAT		1939
Qy	1601	PNTEFI--LDTNE-IISGNSIQOFE----LICDKNKIQPYFINFKIKETSYTLVY--		1650
Db	1940	TQIEYISTADVFSSENTALSICAIYPLAIA-----IVPITI-FSHEVKNVAVYVQ		1993
Qy	1651	GNRQNLIVPSYHLD-----SGNISSTVINFSQKLYG----IDRYVNVKVII		1694
Db	1994	INERHKLWLEAEKYLDNGSAKVLISINKATGIIDLSSNNQVLGNLYLDMRENPPIL		2047

RESULT 12

US-09-134-001C-3159	Sequence 3159, Application US/09134001C
Patent No. 6380370	
GENERAL INFORMATION:	
APPLICANT: Lynn Doucette-Stamm et al.	
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS	
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS	
FILE REFERENCE: GPC-007	
CURRENT APPLICATION NUMBER: US/09/134,001C	
CURRENT FILING DATE: 1998-08-13	
PRIOR APPLICATION NUMBER: US 60/064,964	
PRIOR FILING DATE: 1997-11-08	
PRIOR APPLICATION NUMBER: US 60/055,779	
PRIOR FILING DATE: 1997-08-14	
NUMBER OF SEQ ID NOS: 5674	
SEQ ID NO 3159	
LENGTH: 10182	
TYPE: PRT	
ORGANISM: Staphylococcus epidermidis	
US-09-134-001C-3159	

```

Query Match      4.0%; Score 348; DB 4; Length 10182;
Best Local Similarity 19.4%; Pred. No.1.9e-11;
Matches 354; Conservative 317; Mismatches 663; Indels 494; Gaps 97;

QY      20  QDEYVAILNALLEYHNMSSESVVEKYLKLDINN-----LTDN-----YLN-- 61
Db      7557 QKD---TILNHI---FSAPTRSQVGKIASAKQLANTMKALRDSIADNNEILQSSKYFNE 7611
QY      62  TYKSGRNKALKKPKELYLTMEVLEKNNSLTPV-----EKNLHFIMTGG 105
Db      7612 SEQONAYNAQVNAKNII-----NDQPTPVMANDEIQSVLNEVKOTKONLH----GD 7659
QY      106 Q--IND-----TAINYNQWKDVNSDYTVKVFYDSN-----AFLINTLKKT 144
Db      7660 QKLANDXTDAQATNALNYLNAQGRNLETQVQ---NSNSRPEQKVQVVLANQLANDAMKK 7716
QY      145 IVSATWNTLESFRENLDNPEFDYKPKRKMEIYYQKQKHFIDYYSQIBEE--NPEF-- 200
Db      7717 LDDALTCN--DAIKQTSYINEDTSQ-----QVNFDB---YTDGRKNIVAEQTNPNMSP 7765
QY      201 -----IIDNIKIYLSNEYSKDLLEANKY-----IESLNKIITANNNGDNRINLEKFADELDV 252
Db      7766 TNINTIADKI-----TEAKNDLHGVLKQAQOQSINTI-----NQMTGLNQAKQOL- 7813
QY      253 RLYNQELVERWNLAAGSDILRI5MLKEDGGVYLDVDILPGIQPDLFKSINKPDSITNTSW 312
Db      7814 ---NQEIQQQTQTRSEVHQVINKAQLNDS-----MNTLRQSIITDEHEVKQTSN 7858
QY      313 EMIKL-----EAIMKYKEIIPGYTSKNPDMLEEVQSRFSALSSKDSKSEIFPLD 364

```

7859	YINETGNOTAYNNAVDRVKQIIINQTSNPTMNP	---EVERATSNVSKISK-DALHGERELN	7911
365	DIKVSPLVFKIAPANNSSV--INQALISLKDSYCSDL	-----VINOIKRYKILN---	411
7916	DNKNSK-----TFAVHLDNLNOA--	--QKEALTHEIQATIVSOVNNIYNKAKALNDMK	7967
412	---DNLNPISINEGDFNTWKIKFSDKLASISNEDNMFM	-----KITNYL	454
7968	KLKDIIVAQDNVRQSNNYINEDSTPQMYND--	---TINHAQSIIDQVANPTMSHDEIENAI	8024
455	KVGFAPDVRSTIN--LSPGPGVYTGAYQDMLLMPKDN	STNIHLLEPELNFPEP-KTKISQLT	512
8025	N-----NIKHAINALDGB-----	HKLOQAKENA--NULLINSLDNLNAPQDAINRLV	8069
513	E-----QEITSLWSFNQA-----	RAKSQF---BEYKKGYPEGALGED	546
8070	NEAOTREKVAEQLOSAQALNDAMKHLRNSIQOSSVRQES	KYINASDAKKEQYNHAVREV	8129
547	DNLDFAQNTVLDKDYVSKILSSMKTRNKEYIHYI	VOLOQDKISYBASCN--LFSKDPYS	6048
8130	ENIINEQHPTLDEIIL-KQLTDGVQNOANND-LNGVELL	DADKQNAHQSIPTLMHLNQAQQ	8187
605	SILYQK-NIBGSTAYYYYVADAEIKDYRIPYOISNKR	NKTKLFIHGKSEFNTDP	663
8188	NALNEKINNNAVTEVAIIGQAKLJHDHAWENLEESIK	KEQVK-----QSSNYINEDSD	8242
664	ANLDVSLSEIETILN-----LAKADISPKYIEIN	-----LLGCNMFYSYIABETYP	712
8243	VQETYDNADVHTVTEILNQTVPNTLSIEDIEHA	NEVNOAKQLRG-----KQKLYQTIIDLA	8298
713	GKLLKIKIKRVSELMPSISODSITVSANOYEVREIN	BEGKR-----EILD	756
8299	DKELSKLDLDTSQSSSISNQIYTKRTRTEVAQAEI	AKASLNHAMKALNKVYKNADKVL	8358
757	HSGKWINKE-----ESIIKDISSEKEYISFN	PENKILVSKYLHELSTLLQEIERN--	807
8359	SSRFINEDQPEKKAYQOAINHVDSIIRHQTNP	EMDPTVINS-ITHELETAQNHLHGDKQ	8416
808	-----ANSSDIDLEKKVMLTECIN-----	VASNIDR-QIVEGREIEAKNLT	848
8417	LAHAQQAANVINGLIHLNVAQREVIMNTNAT	TREKVAKNJDNQAALDKAMETLQOVV	8476
849	SDSINYIKNEFKLIESDSLYDLKHQNGLDSDH	FSIFEDISKTENGFRIFRINKETGNS	908
8477	AHK--NNILNDSKYLN--EDSKYQQOYDRVI	ADAEQLLNTQNTPTLEPYKVDIVKDNV--	8530
909	IFITEKEITSEYATHISKEISINOTIFDNVNG	KLVKVKNLDAAEHVNTLSNAPFIQSL	968
8531	--LANEKILFG--AEKLSYDKSNANDEI-----	KHMNYLNA-----	8563
969	IEYNTTESLSNLSVAMKVQ--VYAQLFSTGLNT	ITDASKVVELVSTAL-----	1015
8564	-QKQSKIDMISHAALRTEVKQLLOQAAILDEAM	KSLDQVY-ITDTLLPNTYEASEDK	8621
1016	DETIDLPTLSEGLPIIATII-DGVSLG--AAIK	ELSETNDPL-----LRQIEIAKI	1064
8622	KEKVD--QTVSHAQAIIDKINGSNVSLDQVR	QALEQLTOASENLGDGQRYE-EAKVHANQ	8678
1065	---GIWAVNULTAATA--IVTSAIGTASGPS	ILLVPLAGISAGIPSLVNNELILOQKATK	1119
8679	TIDQLTHLSLQOQTAKESVKNATKLEE-----	IATVSNNAAQALNKVMGK	8723
1120	VIDYFKHISLAETEGAFITLLDDKIIMPQDD	VLSEIIDFNNNSTILGKEIWRAEGGSGHT	1179
8724	LEQFINHADSVENSDYRQADDDKIITAYDEAL	EHGQDIQKTNATQNETK-----	8772
1180	LTTDIDHFFSSPSITYRKPMLSITYDVLN-IK	KEKIDFSKDLMLPNAPNRVFGYEMGWTP	1238
8773	-----QALQOLIYAEVSLNGFERLNHAR	PRALPEYIKSLEKINNQAQSALEDKVTQSH	8824
1239	GFRSLD---NDGTKLJDRIRDHYEGOFYWR	YAFIADALITKLKPRYEDTNVRINLDGNT	1295
8825	DLLELEBHIHVEGNTLNDIMGE-----	LANAIWN--NYAPTKASIN-----	8862

QY 878 LDDSHFISPEDISKTENGF-RIRFINKETGNSIF-IETEKEIFSYBATHISKEISNIKOT 935
Db 1615 IEGSEPILEERLCVNLNAGSRVTRWTDNCNLMHQNQLEIFDGNVAHISTWLQ-AEA 1673
QY 936 IFDNVNGK-----LVKKV--NLDAH--EVTNLSAFFIOS-----LIEY 971
Db 1674 LLDEIEKPTSKQEEIVKRLVSELDANLQVENVRDQALILMNARGSSRELVEPKLAE 1733
QY 972 NITKESLS-----NLSVAM-----KVQVYQAFLSTCL-----NTITDASKVE--L 1010
Db 1734 NNFEEKVSHIKSAKLLIAQEPYQCLVTFETFGVPSFDEKLENDIENMLKVEKHL 1793
QY 1011 VSTALDETIDLLPTLSEGLPIIATIIDGVSGLAAIKELSETNDPLLROIE-----AKIGI 1066
Db 1794 ESSEDEKMD-----EESAQIEEVLRGEEMLHQPMEDNKEKIRL 1834
QY 1067 MAVNLTAASTAIVTSAL-----GIASGPSILLVPL-----AGISAGIPS 1105
Db 1835 QULLLHTRYNKIKAIPIQQRKMGQLASGRSSLLPTDYLVEINKILLCMDDVELSLNVPE 1894
QY 1106 LVNNELILOKATKVIDYFKHISLAETEGAFILLDDKIIM---PODDLVLSEIDFNNSI 1162
Db 1895 L--NTAIYEDFS-----FQEDSLKNIKDQDLKGLGEQIAVIERKQPDVIL----- 1936
QY 1163 TLGKCEIWRAGGSG---HTLT-----DDIDHFPSPSITY---RKFWLSIYDVNLK 1209
Db 1937 -----EASGPEAIQIRDTLTQLNKWDNRINRMSDRKGFCDRAWEWRQFHCOLDNL 1988
QY 1210 KKKIDPSKOLMVLNPN-----RVFGYEM-----GWTGFRSLDNDGKLLDRIND 1256
Db 1989 TQWITEABELLVDTAPGGSLOLEKARIHQOELEVGISHSQPSAALNRTGDGIVQKL-S 2047
QY 1257 HVEGQPYWYFAFIA---DALITKLPRYEDTVNRINLDCNTRSFIVPVITTEQIRKNLS 1313
Db 2048 QADGSKPEKLKLAGLNQRDAIAEVRKDR-----QPLRGESK-----QVWK--- 2088
QY 1314 YSFYSGSGSYSLSPYNNIDNLNVDNVTWVIDVNVV-KNITIE-SDEIQKGLIENI 1371
Db 2089 -----YRQLD-----EICWLTKAHAKQKSTTELGENLOE---LRDL 2125
QY 1372 LSKLNIEDNKI-ILNHNHTINFYGD-----INESNRFLSTFS-ILEDI-----NI 1414
Db 2126 TQEMEVAEKLKWLNRTELEMLSDKSLSPERDKISESLRTVMNTWINKICREVPITLKEC 2185
QY 1415 IIEIDLVSYSKILLSGCNMK--LIENSSDI 1443
Db 2186 IQEPSSVSQT-RIAAHPNVQKVLVSSASDI 2215

RESULT 15

US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 3.2%; Score 281; DB 4; Length 3696;
Best Local Similarity 18.0%; Pred. No. 5.1e-08;

Matches 335; Conservative 284; Mismatches 693; Indels 546; Gaps 79;
QY 48 KKDINNLNTYLVATYKSGRNKAL--KKPKLEYLTMEVLELKNLSLTPVKNLHFIIWIGG 105
Db 513 KLRPN-----ISTPREVTNEAIKYTFSEY-----YINTNDNTVT-----G 550
QY 106 QINDTAINYQWQDVNSDYTVKVFYDSNAFL-----INTLK-----KTIVESATNNLTESF 157
Db 551 QOTFPSINVNN-KDDLSEQVNDIIPSNYTLASVKNYKLERAQTVLDBETNNT--PF 607
QY 158 RENLNDPFD-----YKPYRKRMEIYYQKHFIIDYKYSQIE--ENPEFIIDN 204
Db 608 NORYSQTQIDLLHELQTLINRVASRE--INDKAQEMTDVAVDSTELTTEEKOTLVDQ 665
QY 205 I--IKTYSLSYKDL-----EALNKYIEESLNKITANNNDIRNLEKFADELDV 252
Db 666 IENHKEISNNIDDELTDGVERVKEAGLTLESDTPHPVTKPNARQVNNRADOQKTLI 725
QY 253 RLYNOELVERWNLAA-----ASDIL-----RISMLKEDGGVYLDVILPGIOPD 296
Db 726 RNNHEATTEEQNEAIRQVEAHSSDAIAKIGAEATDTTVNEARDNGTKLIATDV-----PN 780
QY 297 LFKSINKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEBQVRSFSAKSKSDK 356
Db 781 PTKAARAAAVTNGANSKIK-----LSGPGVYTGAYQDLLMFKD-----NSTN 490
QY 357 SEIFLPDDIKVSPLEVKIAFANNVINOALISLSDYCSDLVINQIKRYKILNDLNP 416
Db 816 -----AIALVNRK-----KDEAIQN--INTAOG-----NDDVTE 842
QY 417 SINEGTDFTNTWKI-----FSDKLASI--SNEDNMFMKITYLVKGVFADPVRSTIN-- 467
Db 843 AQNNGT--NTIQVPLTPVKQNAIATINAKADEQKRLIQANNATTBEKADAERKVEA 900
QY 468 -----LSGPGVYTGAYQDLLMFKD-----NSTN 490
Db 901 VITANQNTTATNRDQVDTGSGIISAIISPATKIKEDARAAVEAKAIQNOQINSN 960
QY 491 IHLLEPELRNFPFKTKISQLTQEITSLWSFNQARAKSQFEYKKGFEAGLGEDDNL 550
Db 961 MATTEEK-----EDALNQVEAHKQAAIATINQAQSTQOVSEAKNGNINTINQOPNAV 1013
QY 551 FAQN--TVLQDYVSKIL-----SSMKTRNKYIHYIVQLQGDKISYEASCNLFSDOPY 603
Db 1014 KNWTKTILEQKNEKKSIAIAQTPDATTEEQEAVSAVSQAVTNGITHINQAN--SNDV 1071
QY 604 SSILYQK-----NIEGETAYYYVADAETKEI-----DKYRIPYQISNK 643
Db 1072 DQELSNAEQIITHTVNVQKQPQARQALIAKTNEKQSAINSNDEGTIEBEKQAIQSLNDA 1131
QY 644 RNKLTFFCHGKSEBNTDTFANLQVDSLSSSEIETILNLAKAD-----ISPKYIEIN 694
Db 1132 KNLADEQITQAASQNVQNDNALNIGISNT-SKIQT--NFTKKQOARDQVQKQFEKAEALN 1188
QY 695 LLGCMFYSYVAEETYPGKLLKIKDRVSELMPISQ-----DSITVSA-----NQ 741
Db 1189 -----STPATQDEKQDALTRLTOAKETALNDINQAQNTQNVDTALTSGIQNTQ 1240
QY 742 YEVRINEEGKREILD-----HSGKWINKESIIKD-----ISKEYISNP 782
Db 1241 VNVKQKQAKTINDIVQOHKQSIQNDDATTTEEKEVANLNVASQQNVISKIDATTNN 1300
QY 783 KENKIIVKSKYLHELSTLLOEIRNANNSSDIDL---EKKVMLTECEINVASNIDQIVEG 839
Db 1301 QIDGIVSGROSINAITPDTSIKRNA-KNDIDIKAAKKIKIQ--RINDATDEETQENR 1357
QY 840 RIEEAK-----NLTSDSINYIK-NEFKLIESI-----SDSLYDLKHQGLDSDH 882
Db 1358 KIEEAKIEAKONIQRNSTQDVNEAKTNGINKIENTPATTVKSEARQAVQNKANEQINH 1417
QY 883 FISFEDISKTENGFPIRINKETGNSIFIEETEKEIFSYBATHISKEISNIKOTIFDNVNG 942
Db 1418 IQNTPDATNEEKQEAIRNVSAE-----LARVOAQINAEHTTQGVKTIKDDAITSLSRINA 1472

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 5, 2003, 19:27:13 ; Search time 23.125 Seconds
(without alignments)
7069.698 Million cell updates/sec

Title: US-09-126-816B-6_COPY_1_1700
Perfect score: 8677
Sequence: 1 MNLVNAQLQKVVYKFRIO.....YLYGIDRVYKVIAPNLYT 1700

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8646	99.6	2364	2 I40884	Cytotoxin L - Clostrid
2	6733.5	77.6	2366	2 S10317	toxin B - Clostrid
3	6711	77.3	2367	2 S70172	toxin A - Clostrid
4	4458	51.4	2710	2 A37052	alpha-toxin - Clostrid
5	2108.5	24.3	2178	2 S55805	toxin B - Escherich
6	496.5	5.7	3169	2 T00296	rhoctry protein -
7	470.5	5.4	2401	2 T28676	rhoctry protein -
8	437	5.0	2269	2 T28677	hypothetical prote
9	424	4.9	2166	2 G70163	hypothetical prote
10	423.5	4.9	4688	2 F82885	adherence factor T
11	403	4.6	3225	2 D81702	hypothetical prote
12	398	4.6	1979	2 C71622	hypothetical prote
13	390	4.5	3724	2 T18427	hypothetical prote
14	386.5	4.5	3394	2 T18501	hypothetical prote
15	370	4.3	1711	2 T18429	hypothetical prote
16	359	4.1	1939	2 T18372	repeat organellar
17	356.5	4.1	4981	2 T18489	reticulocyte-bind
18	351	4.0	2829	2 A42771	adherence factor T
19	348.5	4.0	3335	2 H81702	ORF MSV156 hypothe
20	345.5	4.0	1127	2 T28317	conserved hypothet
21	342	3.9	2819	2 A90551	hypothetical prote
22	340	3.9	5005	2 F82884	serine/threonine-s
23	337	3.9	2485	1 H71621	asparagine/asparta
24	330.5	3.8	3844	2 T18402	adherence factor T
25	330	3.8	3255	2 G81702	hypothetical prote
26	327.5	3.8	4550	2 T18440	DNA-directed RNA p
27	327	3.8	2910	2 T28156	major merozoite su
28	318.5	3.7	1701	2 A26868	hypothetical prote
29	318	3.7	3973	2 B71612	

30	317.5	3.7	1726	1 SAZQGM	major merozoite su
31	317.5	3.7	1726	2 A45948	major merozoite su
32	317.5	3.7	2599	2 F90608	ABC transporter pe
33	316.5	3.6	1701	2 A54498	major merozoite su
34	316	3.6	1447	2 F82909	hypothetical prote
35	314.5	3.6	2496	2 A71616	secreted protein P
36	314	3.6	1819	2 D97033	uncharacterized pr
37	313	3.6	1306	2 T28313	ORF MSV152 probabl
38	311	3.6	1302	1 JC6009	surface-located me
39	310.5	3.6	1830	2 E82909	conserved hypothet
40	309.5	3.6	2244	2 F90563	hypothetical prote
41	309	3.6	1639	2 S05603	major merozoite su
42	306.5	3.5	1516	2 E71619	RAD2 endonuclease
43	306.5	3.5	2136	2 A05037	hypothetical prote
44	305.5	3.5	1957	2 T38077	hypothetical coile
45	303.5	3.5	1802	2 G71616	hypothetical prote

ALIGNMENTS

RESULT 1

I40884
cytotoxin L - Clostridium sordellii
C:Species: Clostridium sordellii
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40884
R:Green, G.A.; Schue, V.; Montell, H.
Gene 161, 57-61, 1995
A:Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium 80
A:Reference number: I40884; MUID:95369733; PMID:7642137
A:Accession: I40884
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2364 <RES>
A:Cross-references: EMBL:X82638; NID:g1000694; PIDN:CAA57959.1; PID:g1000695
C:Superfamily: cpl repeat homology
C:Keywords: cytotoxin

Query Match 99.6%; Score 8646; DB 2; Length 2364;
Best Local Similarity 99.8%; Pred. No. 1.2e-303;
Matches 1696; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	1	MNLVNAQLQKVVYKFRIOQEDEVVAILNALLEEYHNMSSESVVEKYLKLDKINNLTNYL	60
Db	1	MNLVNAQLQKVVYKFRIOQEDEVVAILNALLEEYHNMSSESVVEKYLKLDKINNLTNYL	60
Qy	61	NTYKSGRNKALKKPKYLTWEVLEKNSLTPVEKNLHFIWIGQINDTAINYNQWKD	120
Db	61	NTYKSGRNKALKKPKYLTWEVLEKNSLTPVEKNLHFIWIGQINDTAINYNQWKD	120
Qy	121	VNSDYTVKVVDSNAFLNTLTKTIVESATNTLTSPFRENLDNDFDYNKFKRMEIY	180
Db	121	VNSDYTVKVVDSNAFLNTLTKTIVESATNTLTSPFRENLDNDFDYNKFKRMEIY	180
Qy	181	DKQKHFDIYKSOIQEENPEFIIDNIIKTYLSNEYSKOLEALNKYIEESLNKITTANNNDI	240
Db	181	DKQKHFDIYKSOIQEENPEFIIDNIIKTYLSNEYSKOLEALNKYIEESLNKITTANNNDI	240
Qy	241	RNLEKFADEDLVRVYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDLPFKS	300
Db	241	RNLEKFADEDLVRVYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDLPFKS	300
Qy	301	INKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSSKSKSEIF	360
Db	301	INKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSSKSKSEIF	360
Qy	361	LPLEDDIKVSPLEVKIAPFANNVINQALISLKDSCDLVINOIKRYKILNDNLNPSINE	420
Db	361	LPLEDDIKVSPLEVKIAPFANNVINQALISLKDSCDLVINOIKRYKILNDNLNPSINE	420
Qy	421	GTDFNTTKIFSDKLASISNEDNMFMKITTNYLVKGFADPVRSITNLSGPGVYTGAYOD	480
Db	421	GTDFNTTKIFSDKLASISNEDNMFMKITTNYLVKGFADPVRSITNLSGPGVYTGAYOD	480

```
Db 421 GTDFNTMTKIFSDKCLASISNEDNMFMKIITNLYKVGFAFAPVRSTINLSGPGVYTCAYQD 480
Qy 481 LLMFKDNSTNIHLPELRFNFPKTKISQLTEQETISLWSFNQARAKSQPEYKKGVE 540
Db 481 LLMFKDNSTNIHLPELRFNFPKTKISQLTEQETISLWSFNQARAKSQPEYKKGVE 540
Qy 541 GALGEDNDLFAQNTVLDKDYVSKKILSSMKTRNKEYIHYIVQLQGDKISYEASCNLFK 600
Db 541 GALGEDNDLFAQNTVLDKDYVSKKILSSMKTRNKEYIHYIVQLQGDKISYEASCNLFK 600
Qy 601 DYPSSILYOKNIEGSETAYVYVADAEIKEDIKYRIPYOISNKRNIKLTFFIGHGSEFNT 660
Db 601 DYPSSILYOKNIEGSETAYVYVADAEIKEDIKYRIPYOISNKRNIKLTFFIGHGSEFNT 660
Qy 661 DTFANLDVDSLASETETILNLAADISPKYIEINLLGCNMFPSYISABETYPCKLLKIK 720
Db 661 DTFANLDVDSLASETETILNLAADISPKYIEINLLGCNMFPSYISABETYPCKLLKIK 720
Qy 721 DRVSELMPSISQDSITVSANQYEVRIENERGKEIILDHSGKWKINKESIIKOISSKEYISF 780
Db 721 DRVSELMPSISQDSITVSANQYEVRIENERGKEIILDHSGKWKINKESIIKOISSKEYISF 780
Qy 781 NPKENKIIVSKYLHELSTLLQEIERNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840
Db 781 NPKENKIIVSKYLHELSTLLQEIERNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840
Qy 841 IEBAKNLTSDSINYIKNEPKLIESDSLYDLKHQGLDSDSHFISFEDISKTEGFRIRF 900
Db 841 IEBAKNLTSDSINYIKNEPKLIESDSLYDLKHQGLDSDSHFISFEDISKTEGFRIRF 900
Qy 901 INKETGNSIFITEKEIFSEYATHISKEISNICKDTIFDNVNGKLVKKNVLDAAHEVNTLN 960
Db 901 INKETGNSIFITEKEIFSEYATHISKEISNICKDTIFDNVNGKLVKKNVLDAAHEVNTLN 960
Qy 961 SAFFIQSLLEYNTTKESLNSLVAMKVQVYAOQFSTGLNTITDASKVELVSTALDETID 1020
Db 961 SAFFIQSLLEYNTTKESLNSLVAMKVQVYAOQFSTGLNTITDASKVELVSTALDETID 1020
Qy 1021 LAPTISEGPIIATIIDGVSIGNAIKELSETNDPLLRQIEAKIGIMAVNLTAASATAVT 1080
Db 1021 LAPTISEGPIIATIIDGVSIGNAIKELSETNDPLLRQIEAKIGIMAVNLTAASATAVT 1080
Qy 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELILQKATKVIDYFKHISLAETEGAFITLD 1140
Db 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELILQKATKVIDYFKHISLAETEGAFITLD 1140
Qy 1141 DKIIIMPQDLVISEIDFNNSITLGKCEIWRAGSGGHTLTDDIDHFFSSPSITVRKPKWL 1200
Db 1141 DKIIIMPQDLVISEIDFNNSITLGKCEIWRAGSGGHTLTDDIDHFFSSPSITVRKPKWL 1200
Qy 1201 SIYDVNLIIKKEIKIDFSKDLMLPNAPNRVFGYEMGWTGFRSLDNDGTKLLDIRDHVEG 1260
Db 1201 SIYDVNLIIKKEIKIDFSKDLMLPNAPNRVFGYEMGWTGFRSLDNDGTKLLDIRDHVEG 1260
Qy 1261 QFYWRFYAFIADALITKLAPREDTNRINLGNTRSFIVPVITTEQIRKNLSYFYSGS 1320
Db 1261 QFYWRFYAFIADALITKLAPREDTNRINLGNTRSFIVPVITTEQIRKNLSYFYSGS 1320
Qy 1321 GSYSLSLSPYNNMIDLNVENDTWIDVDNVVKNITTESDEIQKGEIENILSKLNIEDN 1380
Db 1321 GSYSLSLSPYNNMIDLNVENDTWIDVDNVVKNITTESDEIQKGEIENILSKLNIEDN 1380
Qy 1381 KIILNNHTINFVGDINESNRFSLTFSIILEDINIIIEIDLVSKEYILLSGCMKLIENS 1440
Db 1381 KIILNNHTINFVGDINESNRFSLTFSIILEDINIIIEIDLVSKEYILLSGCMKLIENS 1440
Qy 1441 SDIQQKIDHIGFNGEHQKIFYSIDNETKYNGFIDYSKKEGLFAEFSNESIIRNIYMP 1500
Db 1441 SDIQQKIDHIGFNGEHQKIFYSIDNETKYNGFIDYSKKEGLFAEFSNESIIRNIYMP 1500
Qy 1501 DSNLPIYSSKOLKDIRINKGDKVLLIGNYFKDDMKVSLSTFIETDNTIKLVGVYLDN 1560
Db 1501 DSNLPIYSSKOLKDIRINKGDKVLLIGNYFKDDMKVSLSTFIETDNTIKLVGVYLDN 1560
```

```
Qy 1561 GVAQILKPMNNAKALNTSNLMNLFESINIKNIFYNNLDNPIBILDTNFIISGNSIG 1620
Db 1561 GVAQILKPMNNAKALNTSNLMNLFESINIKNIFYNNLDNPIBILDTNFIISGNSIG 1620
Qy 1621 QFELICDKDKNIQYFNFINKIKETSYTLVGNRQNLIVPEPSYHLDDSGNISSTVINFSQK 1680
Db 1621 QFELICDKDKNIQYFNFINKIKETSYTLVGNRQNLIVPEPSYHLDDSGNISSTVINFSQK 1680
Qy 1681 YLYGIDRYVNVKVIAPNLYT 1700
Db 1681 YLYGIDRYVNVKVIAPNLYT 1700

RESULT 2
S10317
toxin B - Clostridium difficile
C:Species: Clostridium difficile
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S10317; S21894; S22434
R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.
Nucleic Acids Res. 18, 4004, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.
A:Reference number: S10317; MUID:90326540; PMID:2374729
A:Accession: S10317
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-2366 <BAR>
A:Cross-references: EMBL:X5138; NID:g40442; PIDN:CAA37298.1; PID:g40443
R:Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, N.;
submitted to the EMBL Data Library, July 1991
A:Description: Comparative analysis of Clostridium difficile toxins A and B.
A:Reference number: S21894
A:Accession: S21894
A:Molecule type: DNA
A:Residues: 1271-2366 <EIC>
A:Cross-references: EMBL:X60984; NID:g40445; PIDN:CAA43299.1; PID:g40446
R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn,
Mol. Gen. Genet. 233, 260-268, 1992
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.
A:Reference number: S22434; MUID:92293124; PMID:1603068
A:Accession: S22434
A:Molecule type: DNA
A:Residues: 1791-2366 <VON>
A:Cross-references: EMBL:X60984
C:Genetics:
C:Gene: toxB
C:Superfamily: cpl repeat homology
C:Keywords: cytotoxin
```

```
Query Match 77.6%; Score 6733.5; DB 2; Length 2366;
Best local Similarity 76.4%; Pred. No. 8.4e-235;
Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;

Qy 1 MNLVNAQLOKQWVVKPRIQDEYVAIINALAEYHNNESSVWEKYLKLDNNLTNNVL 60
Db 1 MSLVNRKQLEKXANVRFTQDEYVAIILDALEEVHNSENTVVEKYLKLDNNLTNNVL 60

Qy 61 NTYKSGRNKALKKPKLEYLTWELVELKNNSLTPVEKNLHFHWIGQINDTAINYNQWK 120
Db 61 DTYKSGRNKALKKPKLEYLTWELVELKNNSLTPVEKNLHFHWIGQINDTAINYNQWK 120

Qy 121 VNSDYTVKVFVDSNAFLINTLKKTIVESATNNTLESFRENLDNDFDYNKPKRKMELIY 180
Db 121 VNSDYNNVNFVDSNAFLINTLKKTIVESAINDTLESFRENLDNDFDYNKPKRKMELIY 180

Qy 181 DKQHFIDYKSGQTEENPEFIIDNIIKTYLSNEYSKOLEALNKYTEESINKITANGNDI 240
Db 181 DKQKNFINYKQAEENPELIIDDIIVKTYLSNEYSKDELDENLYTEESINKITQNSGNDV 240

Qy 241 RNLEKFADEDLVRLYNQVELVERWNLAAASDILRISMLKDEGGVYLDVILPGIQDLPKS 300
Db 241 RNFBEFKNGESFNLYEQELVERWNLAAASDILRISALKEIGWYLDVDMPLPGIQDLPKS 300
```


A:Molecule type: DNA
A:Residues: 1-55,'V',57-2079,'L',2081-2549,'S',2551-2710 <SAU>
A:Cross-references: EMBL:X51797

C:Genetics:

A:Gene: toxA

C:Superfamily: Clostridium difficile toxin A; cpl repeat homology

C:Keywords: cytotoxin; enterotoxin

F:1820-1839/Domain: cpl repeat homology <CP01>
F:1840-1860/Domain: cpl repeat homology <CP02>
F:1861-1881/Domain: cpl repeat homology <CP03>
F:1933-1952/Domain: cpl repeat homology <CP04>
F:1953-1973/Domain: cpl repeat homology <CP05>
F:1974-1994/Domain: cpl repeat homology <CP06>
F:1995-2015/Domain: cpl repeat homology <CP07>
F:2067-2086/Domain: cpl repeat homology <CP08>
F:2087-2107/Domain: cpl repeat homology <CP09>
F:2108-2128/Domain: cpl repeat homology <CP10>
F:2129-2149/Domain: cpl repeat homology <CP11>
F:2201-2220/Domain: cpl repeat homology <CP12>
F:2221-2241/Domain: cpl repeat homology <CP13>
F:2242-2261/Domain: cpl repeat homology <CP14>
F:2315-2334/Domain: cpl repeat homology <CP15>
F:2335-2355/Domain: cpl repeat homology <CP16>
F:2356-2376/Domain: cpl repeat homology <CP17>
F:2377-2397/Domain: cpl repeat homology <CP18>
F:2449-2468/Domain: cpl repeat homology <CP19>
F:2469-2489/Domain: cpl repeat homology <CP20>
F:2490-2510/Domain: cpl repeat homology <CP21>
F:2562-2581/Domain: cpl repeat homology <CP22>
F:2582-2602/Domain: cpl repeat homology <CP23>
F:2653-2672/Domain: cpl repeat homology <CP24>
F:2673-2694/Domain: cpl repeat homology <CP25>

Query Match 51.4%; Score 4458; DB 2; Length 2710;
Best Local Similarity 50.7%; Pred. No. 8.4e-153;
Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;

QY 1 MNLVKAOLQKMYVYKFRIOEDYVAILNALDEYHNMSSVVEKYKLKIDNNLTNDYL 60
DB 1 MSLISKEELIKLAY-SIRPRENEYKTLTNDYENKLTNNENKYLQKLKLNESIDVPM 59
QY 61 NTYKSGRNKALKPKLEYLTMVEBLEKNNSLTPVEKNLHFVIGGQINDTAINYNQWKD 120
DB 60 NKYTSRRNALSNLKKDILKEVILKNSNTSPVEKNLHFVIGGVSVDIALEYIKQWAD 119
QY 121 VNSDTYKPYVDSNAFLINTLTKTIVESATNTLTESFRENLDNDFEDYKFRKMEIY 180
DB 120 INAYNIKLYDSEAFVNLTKKAIVESSTTALQLEBEEIQNPQPDNNKPKYKRMETIY 179
QY 181 DKQKHFIDYKSOIBENPEFIIDNIKTLYSNEYSKDLALNKYTBESLNKITTANNNDI 240
DB 180 DKQKHFINYKSOINKPVTPTDDIYKSHLVSEYRDETVELSYRTNSLRKINSNHGDI 239
QY 241 RNLEKFAEDLVRLYNQELVERMNLAAASDIIRISMLKBDGGVLDVILPGIQDPLFKS 300
DB 240 RANSLFTEQELNLYSQELLNRGNLAAASDIIRLLALKNFGGVLDVDMPLGHSDFPKT 299
QY 301 INKPSITNTNEMIKLEIMKYKEYIPGYTSKNFMDLDEEVORFESALSSEKSEIF 360
DB 300 ISRPSIGLDRWEMIKLEIMKYKYYINNTSENFEDKLDQQLKDFKLIIESKSEIF 359
QY 361 LPLDDIKVSPLEKTAFAFANNVINOALISLKDYSCLVINQIKRYKITLNDNLAPSINE 420
DB 360 SKLENLVNLDLBIKIAFALGVSINOALISKQGSYLTNLVIEQVKRYQFNLNOLNAPAES 419
QY 421 GTDFNTMTKIFSDKLASINEDNMFMKIKITNLYKGFAPDVRSTINISGPGVYTCAYOD 480
DB 420 DNNFTDTTKIFHDSLFNSATAENSHFLTAKIAPYLVGFMPPEARSTISLGPAYASAYD 479
QY 481 LLMFKDKNSTNIHLLPELNRNPFPPKTKISQLEQEITSLWSFNQARAKSQFEYKKGYPE 540
DB 480 FINLOENTIEKTLKASDLTEFKFPENNLSQLTEQEINLSWSFDQASAKYQFEKYVRDYG 539

QY 541 GALGEDDNLDAFQONTVLDDKQY-VSKKILSS--MKTRNKEYIHYIYVQLQGDKIYSBASCNL 597
DB 540 GSLSEDNGVDNKNTALDKYLLNNKIPSNVNBAGSKNYVHYIIQLQGGDDISYEATCNL 599
QY 598 FSKDPYSSILYQKNIEGSETAYYYIVAD--AEIKEDIKRIPYOISNRKNIKLTFIGHGK 655
DB 600 FSKNPKNSIIIIQRNM--NESAKSYPLSDGSGSILELNKYRIPERLKNKEKVKVTFIGHGK 657
QY 656 SEFNTDTFANLDVSLSLSEIETILNLAKADISPKYIEINLLGCNMFYSYIAVEEYPGKL 715
DB 658 DEFNTSEFARLSVSLSNEISSFLDTIKLDISPKRVEVNLGCNMFSDVFNVEEYPGKL 717
QY 716 LLKIKORVSELMPSIQSDSIIVSANQYEVRIINEGRBILDHSGKWINKEESIIKDISK 775
DB 718 LLSIMDKITSTLPDVNKNISITIGANQYEVRIINSEGRKELLAHSGKWINKEEAIMSDLSK 777
QY 776 EYISFNPKENKILVSKYLHELSTLLOEIRNANSSDDILEKKVMLTCEINVAENIDRQ 835
DB 778 EYIPFDSIDNKLKAKSKNIPGLASISEDIKTLLLDASVSPDTKFTLNNLKLNISSIGDY 837
QY 836 IVEGRIEBAKNLTSDSINVIKNEPKLIBSISDSLYDLKHQGLDDSHFTSPEDISKTENG 895
DB 838 IYYEKLEPVKLIHNSIDDLDEFNLLNENSDLEYELKLNLDSEKYLISFEDISEKNST 897
QY 896 FRIRFINKETGNSIPFIEKEIFSEYATHISKEISNIKDTIPDNVNGKLKVKVNLDAHE 955
DB 898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLLDNIQLDHTSQ 957
QY 956 VNTLSAFFIOSLIEYNTTKSLSNLSVAMKVQVYLAQFSTGLNTITDASKVVELVSTAL 1015
DB 958 VNTLNAAPFIOSLIDYSSNKDVLNLSVSKVQVLAQFSTGLNTIYDSIQVNLNISNAV 1017
QY 1016 DETDLLPLTSLGPIIATIIDGVSLGAIKELSETNDPLLRQETAEKIGIMAVNLTAAS 1075
DB 1018 NDTINVLTPTTETGPIVSTILDGINLGAAILKELDEHDPLLKKEBAKVGLVAINMSUSI 1077
QY 1076 TAITVSALGIASGFSILLVPLAGISAGIPSLVNNELIILQDKATKVIDYFKHISLAETEGA 1135
DB 1078 AATVASIVGIGAEVITIFLLPIAGISAGIPSLVNNELIILHDKATSVVNFNHLSESKYGP 1137
QY 1136 FTLLDDKIIMPODDLVLSEIDFNNNSITLGCETWRAEGSGSHTUTDDIDHFPSSPSITY 1195
DB 1138 LKTEDDKILVPIDDLVISEIDFNNNSIKLGTCTNILAMEGSGSHTVTGNIHDFPSSPSISS 1197
QY 1196 RKPWLSIYDVLENIKEKIDFSKOLAVLPNAPRVFVEMGTWPPRSLDNDGCTKLLDRIR 1255
DB 1198 HIPSLUSIYSAIGIETENLDFSKKIMLPNAPSRVFWMETGAVPGURSLDNDGCTKLLDRIR 1257
QY 1256 DHYEQFYRYFAFTADALITKLKPRYEDTVNRINLDGNTRSFIYVPIITTEQIRKNLSYS 1315
DB 1258 DLYPKGFYRFAVFP-DYAITTLKPVYEDTNIKIKLDKDRNFIMPTITTTNEIRNKL SYS 1316
QY 1316 FYGSGGSYLSLSPYNNMIDNLVENDTWIDVNVNKNITTESDEIOKGEIENILSKL 1375
DB 1317 FDGAGGTSLLSYSPISNTINLSKDDLWIFNIDNEVREISIENTGTIKGKLIKOVLSKI 1376
QY 1376 NIEDNKIILNKHNTNFYGDINESRFSISLTSILEDDINIILDLVSKSYKILLSCNCK 1435
DB 1377 DINNNKLIIGQITIDFSGDIDNKKORYIFLTCELDKIKLILIEINLAKSYSLLSGDKNY 1436
QY 1436 LIENSDDIQQKIDHIGFNGEHOKYIFYSYVID-NETKYNGFIDYSKKEGLFTAFBNESII 1494
DB 1437 LISNLSTIEKINTLGLD---SKNTAYNTDESNNKYFGAI-----SKTSQKSII 1483
QY 1495 RNIYMPDSNNL-----FIYSSKDL--KOIRIINKGDVKLLIGNYFKD---DMKVSUS 1541
DB 1484 H--YKKDSKNILEFYNDSTLEPNSKDFTAEDINVMFKDIDINTIGKYVYDNDNTKSIDFS 1541
QY 1542 FTIEDNTIKLVLDENGVAQIILKFMNNAKSALNTSNLMNPLESINIKNIFYNLDP 1601
DB 1542 ISLVSKNQVKNGLYNLESVSYSLDFVKNSDGHENTFMNMLFLDNISFWKLFGPE--- 1598
QY 1602 NIEFILDNTNFIISGNSISQFELICDQKQNIQPYFINPKIKETSVTLVYVGNRQNLIVBPS 1661

Db 1599 NINFDVYKFTLVGKTNLGVVEFCIDNNKNDIDYFGWKTSSTSSKSTIPSGNGRNVVBEI 1658
 Qy 1662 YHLDSCNLSSTVINFSQKYLVDIBRYNKNVLIAPNLYT 1700
 Db 1659 YN-PDTGEDISTSLDSYBPLYGIDRYINKVLIAPDLYT 1696

RESULT 5
 S55805
 alpha-toxin - Clostridium novyi (ATCC 19402)
 C:Species: Clostridium novyi
 A:Variety: ATCC 19402
 C:Date: 28-Oct-1996 #sequence revision 08-Nov-1996 #text change 05-Nov-1999
 C:Accession: S55805; S71294; S71158; S44273; I40834; S44272
 R:Hofmann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.
 Mol. Gen. Genet. 247, 670-679, 1995
 A:Title: Sequencing and analysis of the gene encoding the alpha-toxin of Clostridium novyi
 A:Reference number: I40834; MUID:95342160; PMID:7616958
 A:Accession: S55805
 A>Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-2178 <HOF>
 A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA88565.1; PID:g755724
 A:Accession: S71294
 A:Molecule type: protein
 A:Residues: 1-15 <HOF>
 R:Hofmann, F.
 submitted to the EMBL Data Library, March 1995
 A:Reference number: S71158
 A:Accession: S71158
 A:Molecule type: DNA
 A:Residues: 1-1179, 'LKV', 1183, 'LVTHIGE', 1191-2178 <HOS>
 A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA88565.1; PID:g755724
 R:Hofmann, F.; Habermann, E.; von Eichel-Streiber, C.
 submitted to the EMBL Data Library, July 1993
 A:Description: Sequence analysis of Clostridium novyi alpha-toxin: a member of the family
 A:Reference number: S44272
 A:Accession: S44273
 A:Molecule type: DNA
 A:Residues: 1-243; 1204-2178 <HOF>
 A:Cross-references: EMBL:Z23281
 C:Genetics:
 A:Gene: tcn-alpha
 C:Superfamily: cpl repeat homology
 C:Keywords: virulence factor
 F:1880-1899/Domain: cpl repeat homology <CP12>

Query Match 24.3%; Score 2108.5; DB 2; Length 2178;
 Best Local Similarity 31.2%; Pred. No. 2 5e-68;
 Matches 551; Conservative 358; Mismatches 708; Indels 147; Gaps 46;

Qy 3 LVNKAQLQKVVYKFRIOEDVYVAILNALLEYH-NMSESSVVEKYLKLDINNLTNDYLN 61
 Db 2 LITREQLMKIASIPKRPPEYNLIIDALENPNRDIEGTSVKEIYSKLSKLNELVDNYQT 61
 Qy 62 TYKSGRNLKALKFKKYLWELVLEKNSLTP--VEKNLHFTWIGGQINDTAINYNQWK 119
 Db 62 KYPSSGRNLALFNDRSLYSSELRELKNGRTSTIASKLSFTWIGGPISDQSLEYNNMK 121
 Qy 120 DVNSDYTVKVFVDSNAFLNTLKTIVESATNTLTESFPRENLDPEFDYKFKRMEII 179
 Db 122 MFNKNIRLFDKNSLLVNTLKTALIQESSKVIIEQNSNILDGTGYHKNFYSDRMKUI 181
 Qy 180 YDKQHFIDYKSIENPEFIIDNIKTYSNEYSKOLEALNKYIEESLNKITANNNGND 239
 Db 182 YRYKRELKMLYENMKQNS--VDDIIINFLSNYFYDKLNNQKNNNNKMIAGATD 238
 Qy 240 IRLNLEFAEDLVRNLNOELVERWNLAASDILIRSMLEKGGVYLDVILFGIQDPLPK 299
 Db 239 I-NTENILNTKLKSYYYQELIQTNLAASDILIRAILKKGVCYCDLDFPGVNLNLFN 297
 Qy 300 SINKPDSITNTSWEMIKLEAIMKYEYIPGYTSKNPMDLDEEVQSFESALSSKSKSEI 359

Db 298 DISKPNMGDSNVWEAAIFEAIANEKLMMNYPKYTMEQVPSIEKERILSVFN-HDINDL 356
 Qy 360 FLPLDDIKVSPLEV-----KJAFANNSVINQALISLKDYSCLVINOIKRYKILNDNL 414
 Db 357 ILPLGDIKISQLEILLSLRKAATGKTKFSNFIISNNDSLTLNLLISOLENRYETLNSII 416
 Qy 415 NP--SINEGTD--FNTWMKIFSDKLASISNEDNMFMKITNYLKVGFAPDVRSTINLSG 470
 Db 417 QEKFKICETYDSYNSVSELVETTPKNLSWDGSSFYQOIIGYSSGKPEVNSVTFSG 476
 Qy 471 PGVYTGAQODLLMFKDN-----STNIHLLEPELRNFEFPKTKISQLTEQBITSLWSFNQ 524
 Db 477 PNIYSSATCDTVHFIKNTFDMLLSSQNEIFEAS--NNLYFSK-----THDEFKSSWLLRS 529
 Qy 525 ARAKSQPEYKGYFEGALGDNDLDPAQNTVLDKDYVSKKI--LSSMKTRNKEIHIIV 582
 Db 530 NIAEKEFOKLAKTYIGRTLANTYEDGLNFNKRVRTTSELLKVEEYNSKIYENYDLNML 589
 Qy 583 QLOQDKISYEASCNLFSDPYSSILYQKNIEG-SETAYVYVADAEIKEDKRYPIYOIS 641
 Db 590 QLOQDDISYSAVNVFGKNPKSIL----IQGVDDFANVFYFENGIVOSDINNINLSRPN 645
 Qy 642 NKRNILKTFIGHGKSEFNTDTFANLDDVDSLSSEI-----ETILNLAKADISPKEYEINLL 696
 Db 646 DIKKIKLTLIGHGENVFNPKLFGGKTVNDLYNIIKPKQLHLLEREGVILKNKYLKINIL 705
 Qy 697 GGNMFSYIABETYPGKLLKIKDRVSELMF-SISQDSITVSANQYEVNRNEBKEIL 755
 Db 706 GCMFTPKVDINSTPVGKLFNKIS---RDLOPKGFSKNQLEISANKYAIRINREGREVL 762
 Qy 756 DHSGKWINKERSIIKDISKEYISPNKENKIIVKSKYLHELSTLQIRNNANSSDIDL 815
 Db 763 DYFGKWNSTDLIAEQISKNKVTVWNEVENTLSARVEQKNKVAEPAKDI-----NSIIQT 817
 Qy 816 EKKVMTLCEINVASNDRQIVEGRIEBAKNLTSDSI----NYIKNEFKLIESISDSLYD 871
 Db 818 TNNQELKQLVNTVADLLTLYSELLKEDIPELDNIQIKERIILNEISRLHDFSNIL 877
 Qy 872 LKHQGLDDSHFISFEDISKTEGFRIRFINKETGNSIFIEKEIFSEYATHISKEISN 931
 Db 878 FYQKNISNNMIILFDSIIKEKDYNNVKLANKITGETSVIKTYSLSLWNTNFKYKIVDD 937
 Qy 932 IKDTIFDNVNGKLVKKNLDAHAHVNTLNSAFFIQSLIETVNTTKESLNSLVAMQVYA 991
 Db 938 IKGIIVKQINGEFKKAADFIEQNPSSLNSAMLMQLLDYKPYTEILTNTWNTSLKQVYA 997
 Qy 992 QLFSTGLNTITDASKVVELSTALDETIDLLPTLSEGLPIIATIIDGVSLGAALKELSET 1051
 Db 998 QIFOLSIGAIQEAITEIVTIISDALNANFNILSKLVGSSVASVIIIDGINLIALTELKNV 1057
 Qy 1052 NDPLLRQIEIAKIGIMAVNLTAATAIVTSALG---IASGFSILLVPLAGISAGIPLSVN 1108
 Db 1058 KTNFERKLEIARVGMVSYGFILESSLSGLLGATAVSEILGVI SVPVAGLVGLPSLVN 1117
 Qy 1109 NELLIQDKATVIDYFKHISLAETEGATLDDKTIIMPODDLVLSIEIDFNNSITLGKCE 1168
 Db 1118 NILVGEKYNQLDYFSKPYPIVGNPFS--IQDNIIIPYDDIATIELNFKNYKFKYGA 1176
 Qy 1169 IWRABGGSGHTITDDIDHFFSSPSITYRKPMLSIYDVLNIIKKEIDFSDKDLMLVLPNAPNR 1228
 Db 1177 ISGFEKSGHTVGNIDHVSAPSILDHVIE--LSIYPALKLNDTNLP-KGNVLLPSGLNK 1234
 Qy 1229 VFGYEMGWTPGPRSLDNDGTLLDRIDHY-----EGOFYRWYFAFIADALITKUKPRYE 1283
 Db 1235 VYKPEISAAGANSQEGNGVEVLNIRNYVDNSNGNTKFPWKYEAPP--EYFSYMRVEYF 1293
 Qy 1284 DTNVRINLDGNTRSPIVPVTTEQIRKNLSYFVSGSGSYLSLSPYNNIDNLVENDT 1343
 Db 1294 DTKNVILDNENKTIUIIPVTIDENRKNKISVEILGGQYNVILPWNQTNINIVSNKNDI 1353
 Qy 1344 WVIDVNVVKNITIEBDEIQKGELTENTLSKLNIBDNKIILNNHTINFPYGDINESNRFTS 1403


```

Db 1267 NKAAREFYKPSLNLDFLINDERVNLGLIKKYKYLSEIHRHLTNSFADISIP 1326
Qy 1021 LIPTLSEGLPITATII-----DGVSLGAIAK-----ELSETNDPLL 1056
Db 1327 LLQTI---CPSITIIKTYGHQHTNMTVASVVKPYDFNSLGAINDSKSVDPAL 1383
Qy 1057 RQIE-AKIGIMAVNLTAATAIVTSAL-----GIAGSFTL 1092
Db 1384 HTIVEQAKNLSWDFYNTHASITWTIARQHKSTNIEPHQSLDFDRSKGKGLSL 1443
Qy 1093 LVPLAGISAGISLVNN-----ELIQDKATKVIDYFKHISLAETGAFLLDDKLIIMPOD 1148
Db 1444 YLDTGGYGGYQKLRHNDTASLYQTK-----YNDNLKSLNRDDFLKRTQRIITWSN 1497
Qy 1149 DLVLSIDFNNSITIGKCEIWRABEGSGHTLTDI--DHFFSSPSITVRKPLWSIYDVL 1206
Db 1498 EL-----GNRLKNAQLEVLK---DPILTEGILYQRRISLITTEYHSLAQGIS 1547
Qy 1207 NIKK-----EKIDFSKDLAVLPNAPRVFVEMGWTTPGFRSLDNDGTKLL 1251
Db 1548 SFRVTDNPHGCHDFHSLAQLTFIKNI-----TSNR-----NFSLYGSI--- 1589
Qy 1252 DRIRHY-EGQFYRYFAFIADALITKLPRYEDTNVINLDG---NTRSFIVPVITTEQ 1307
Db 1590 --VKIYFSESLNNWKYIKLPLVQTGSLLRDILVITPEKLSLSTGGSLNINHLVPV----- 1642
Qy 1308 IRKNLSYSP-YSGGGSYLSLSPYNNWIDNLVENDTWIDVNDVNNKNTIESDEIQ----- 1363
Db 1643 -----SFYIDIGV-----INGNRISSEST---DVKNKIRSLKINGDILQHYI 1681
Qy 1364 -----KGELNLTSLNLTEDNKI-----ILN---NTINIFY 1392
Db 1682 NTHYLSEEQTKIKOIVDFLQDNTIKVKLESIDIKPISIQOPLHSLSROKEHVKNLL 1741
Qy 1393 -GDINE-SNRF-----ISLTFSLIEDINI-----IIBI---DLVSKSYKILLSGNCMK 1435
Db 1742 SGLLDFSNKLRQKGLSLKTNVSNVNFESKINSITVEVTDLQRLYRVDIDTRVIG 1801
Qy 1436 LI--ENSSDIQKIDHIGNGEHBQKIFYSYIDNETKYNFIDYSK-----KGLFTAEPFS 1489
Db 1802 LTFKEGINSLEALEHMN-----IDAIMSVIGLVQYARMKNDNISAIHDA 1848
Qy 1490 NE-SIIRNLYMPDSNNLIYSSKDLADIRINKNDYK-LLIGNYFKDKMKVSLSFIEDT 1547
Db 1849 GAVSDIKNIV-----DKFLGGILTLTNNRYNPPGVSGASLEGFTSSGLEVCAAS----- 1897
Qy 1548 NTIKLNGV---YLDENGVAQILKFMNNAKSALNTSNLMNFLESINIKNIFYNNLD----- 1600
Db 1898 ---RMGGTAGRYL--SNVAKVIK-----PLLDI--GINWSLYDSSLNHAKA 1938
Qy 1601 -PNIEFI---LDTNP-IIGSNSIGQFE-----LICDKKNIOPIYFNFKIETSITLYV- 1650
Db 1939 TQIEYISTAIDVFSFSINTALSIGAIVPPLAIA-----IVPITI-FSHEVKNAVAVYN 1992
Qy 1651 ---GNRQNLIVEPSYHLDD-----SGNISSTVINFQKLYG---IDRYVNVKVII 1694
Db 1993 QINERHKULAEKYLDNGSAKVLINKATGIIDLSNNQVLGNIYLDIMRENPPIL 2047

RESULT 7
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C;Species: Plasmodium yoelii
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C;Accession: T28676; A45521
R;Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A;Title: Comparison of two members of a multigene family coding for high-molecular mass
A;Reference number: Z20507; MUID:97077455; PMID:8920022
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA

```

```

A;Residues: 1-2401 <SIN>
A;Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1
R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple c
A;Reference number: A45521; MUID:91101660; PMID:2270106
A;Accession: A45521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 2260-2401 <KEE>
A;Cross-references: GB:M34281

Query Match 5.4%; Score 470.5; DB 2; Length 2401;
Best Local Similarity 19.7%; Pred. No. 2.6e-09;
Matches 409; Conservative 328; Mismatches 671; Indels 673; Gaps 102;

Qy 27 ILNLEEYHNHSSSVVEKY-----LKLKDINLTDNLTNYLYKSGRNLKALRFPK 76
Db 62 LINKVSDGNIIEENDIDNFDLSKPKNFKLLEGLNGVFNKWNKYK---NNKLDQFK 118
Qy 77 EYLTEVLELKNLSLTPVEKNLHFLTWIGQINDTAINYNQWKNVSDYT-VKFVYDSNA 135
Db 119 DTMKKIILII-----IQMNEFKGLNDAMTKLKNEGISQK 153
Qy 136 FLINT-LKKTIVESATNTNTLESFRENLDNPFVYKFKRMEIYD----- 181
Db 154 FVINQIIOKQDPKSTYDEKKEGFSSL---ELAKN-WEKKLEIITELKKNEETVOLDI 209
Qy 182 KQKHFIDYKQIENPEFIIDNIKTYSN-----EY-----SKDLEALNKYI 225
Db 210 KIRELIKQIKDIEBQK---IVNDLKLNLKKIKETIEYIKKAVDLKKEIKDNVYI 266
Qy 226 BESLN---KIT--ANNENDIRN-----LEKFAEDIVRLYNO--ELVERWNLAASDIL 272
Db 267 DELAKEPPYQTKYTEKKNEIYNTIKSDFKIYVGDIEQLYNEMFSVQESNIEHNKT 326
Qy 273 RISMLK-EDGGVYLDV-----IMKYKEYIPGY-----TSKNFDMLD 339
Db 327 EILTLTKIDVYNNIQNMETETVSKHLKNIETNNKLSETILDIKIYGEITNELKTL 386
Qy 304 -----PDSITNTSWEMIKLEA---IMKYKEYIPGY-----TSKNFDMLD 339
Db 387 EDFKPKNKGSLNKIDYAKENVQLNVYKSNILEIKKHYNQDQINIDNIKEAKQNYDFK 446
Qy 340 BEVQSFSAKSSKSDKSEIFLPLDDI----- 366
Db 447 EHMKTIPPNEKMYQKPSIEIKIMKDEFLSKVKNYNDPKVYKVESEHNKFTELTNKIK 506
Qy 367 -KVSPLVKIAPAN-----NSVINQALISLKDSY-----C--SDLVINQI 403
Db 507 TEVSDEEIK-KYENKFNDSKSLINETKKSIEEYQNIINTLKKVDDYIKVCLNTNELIINC 565
Qy 404 KNYKILNDNLNPSINEGTDFTNTWKIFSDKLASI----- 438
Db 566 HNKQTTLADKLNQNIKTIKETNSIDKIVTDKFNILTDKKTETLTKTGLSLNNHESNNK 625
Qy 439 -----SNEDNMFM-----IKITNYLKVGPAPDVRSTINLSGPGV 473
Db 626 ELLTVFYDLKANLGNKNMNYKQFNEKEKAVEDIKKNV-----DINKIVSNIEITI 678
Qy 474 YTGAYQDILLMFKDNSTNIHLLEPELRNPFEPKTKISQLTEQBITSLWSFNQARAKSQPEE 533
Db 679 YTSIYN---INEDTENEIG-----KSIELNTKVLKVKANVTNL-----NEIKE 720
Qy 534 YKGYFEGALGEDDNLDPQNTVL--DKDYVSKKTLSSMKT-----RNKEYIHYVQLQG 586
Db 721 KLKDYDFQDFGKEKNIKYPDENKIKNDITLQKIDKSIETLTKKNS--NHIDEIKG 778
Qy 587 --DKISYASCNLFSKDPYSSILYQKNIEG-----SETAYYYVVADEIKEIDKRYIP 637
Db 779 QIDKLKVPNKTMFNEPKK-----IEKKINIVEKIDKKKNY-----KEIDK--LL 824
Qy 638 YQISNKRNIKLTFF-----IGHGKSEFNTDTFANL----- 666

```

Db 825 NEISKIENDKTSLEKLNINISYGRS-----LGNLFQQIDEEKKAEHTIKAMEAYID 878
Qy 667 DVDSL---SSRETTLNL---AKADISPKYIEINILGCMFYSYIY-----706
Db 879 DLNDNIKKKSQIEKEMINMDIKMDI---HKEMKALNISHDDYKIYHTTSKNHBEKISDI 935
Qy 707 -----AEETYPGKLLKIDRVSELMPSISQDSITVSANOYEVINEECKRIL 755
Db 936 RKNSLKIITDSEESYINDIKKEKLVLE-----SQNNNT-DINQYLSKI--ENIYNIL 987
Qy 756 DHSGRKWNKESIIKIDISKEYISPNKENKIIIVKSKYLHELSTLLQEIARNNANGSDIDL 815
Db 988 K-----LNKIKIIDKV--KEYTDEIKNNK-----KINAELS-----NSEKIIITQL 1027
Qy 816 EKKVWLTCSEINVAENIDQIVGEIEEAKNLTSDSIN-----YIK-----856
Db 1028 KENSLKSCOSKIRKTIIDNNYVSECIKNITLKIYVNEKKNNTYFKNAEYNNQVSLN 1087
Qy 857 -NEFKLIESISDLYDLKHQGLD--DSHFISFEDISKTENGFRIRFINKETG-NSIFIE 912
Db 1088 FNNIEMADTKSQYILNIIKNGNGTNTDYNIKELKHKKSNNYK-----DEAGKNTQSIK 1142
Qy 913 TEKEIFSYATHI-----SKEISNIKOTIFDNNVNGKLVKNLNDAAHEVNT-----958
Db 1143 KKNLEFKEYQEVTVLLNKKYIAVELKNKFDKT-KNYSBQIIEIK--DAHNTFTSQADKS 1199
Qy 959 ---LNSAPPIQSLIENYTKESLNSLVAMKVQVYAOQLFSTGLNTITD-----1003
Db 1200 EKKMNEIRNEQIRIEDEAVAKNNKSKAI-LDQLSVPEFPKIKPLKIKOLRTKSDCLKET 1258
Qy 1004 -----ASKVVELVSTALDETIDLPITLSEGLPIIATIDGVSIGAAIKELSETNDPUL 1056
Db 1259 KDIEIKINLSIDTQETKLIENKVINLTLEKLESIKK--QXKNIEDQKKELDEVNSKI- 1315
Qy 1057 ROEIEAKIGIMAVNLTAATAITVSALGIASGFSILLVPLAGISAGIPSLVNNELILODK 1116
Db 1316 -KNIESNVNQHKNY-----EIGIVEKI-----NEIA--K 1342
Qy 1117 ATKVIDYPKHISLATETAGFTLLDDKIIMPQDDVLSEIDFNNSITIGKKEIWAERGS 1176
Db 1343 ANK-----DQIESQKLIIPITKILSPKANDLEGIDTKN--LGK--YNTENNN 1389
Qy 1177 GH-----TLTDDIDHFSSPS---ITYRKPWLSIYDVLNIIKKEIDFSDKOLMVLNAPNRV 1229
Db 1390 IVEEPIKSYDLITHLETVSKEPIYEQ-----IKNKRIITAQNELLTNIKVNKA 1439
Qy 1230 FQYEMGWTGFRSLDNGDKLLDRIDHYEG-----QFYWRY-----FAFIADALIT 1276
Db 1440 KSY-----LDDIEANEPDRIVTHFKNKLVNDKFTWNEYSKVNKGFDNISNS-IN 1488
Qy 1277 KLPKRYEDTVNRINDGNVTRSFIVPITTE-----QIRKNLSYFYSYG 1320
Db 1489 NVK-KSTDENLLNLNLTQETKEMYANIVSKYYSKYEAENIFINIPKLANSLNIQKSS 1547
Qy 1321 G-----SYLSLSPY---NMNIDNLV-----ENDTWIDVNVKNITIESEIOKG 1365
Db 1548 GIDLFPNIALPLVDQKQKOTLTFIPSPKTSYTKISDYSYNTLLDLKRSQELQK 1607
Qy 1366 E-LIENILSKLMIENKIIILNHNTHINFYGDI-NESNRFISLTFISLEIDINIIEIDLVS 1423
Db 1608 EQOALNLFENRLLHDKVQATNELKDTLSDLNKKEQILNKVYLLHKSNNELNKLSCNSQ 1667
Qy 1424 SY-KILLSGNCMKLJENSDDIOQKIDHIGFNGEHQKIYFYSYIDNETKNGFI-DYSKKE 1481
Db 1668 NYDTTLESKYDKIEKSNYKEKENLGIN-----FDVKAEEQFNNDIKDIEKLE 1719
Qy 1482 GLPTAEFSPNESIIRIYNPDSNNLIYKSKOLKDIRINKGDVLLIGNYFKDKMKVSL 1541
Db 1720 NNY-----KHSEKONTFSENNNIIQSKKKUKELT-----NAFNAEIK----1758
Qy 1542 FTIEDNTIKLNGVLDENGVAQILKFMNNAKSA-LNTSNSLMNFLESINIKNIFYNNLD 1600

Db 1759 -----KIEDKIEKNGL--INKLIETROCKMFLTYKTL-----VETLIKTTDY-----1800
Qy 1601 PNIEFILDNTFIIISGNSIGOFELICDKDKNIQVPFNFKIKETSITLYVGNRQ--NLIV 1658
Db 1801 -----TKFITSATFSKEFLKVIDATSNLNDN--TLQTKYDLNQINKHVASWA 1850
Qy 1659 EPSYHLDDSGNI-----SSTVINFSQKYLIGIDRYNVKV 1692
Db 1851 DAT---NDNNNLIEKEATKTINNLT--LFTIDS--NKI 1884
RESULT 8
T28677
rhoptry protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 Haequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28677; C45521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A:Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A:Reference number: 220508; MUID:95021522; PMID:7935623
A:Accession: T28677
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KEE>
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple clones
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: C45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283
Query Match 5.0%; Score 437; DB 2; Length 2269;
Best Local Similarity 20.2%; Pred. No. 3.9e-08;
Matches 396; Conservative 310; Mismatches 659; Indels 592; Gaps 100;
Qy 31 LEEYHNSSESVKYLKL-KDINNLTNYNTYKSGRNKALKFKFYLTWEV-----83
Db 71 LETITELTKSN--BETVLEKEIRELFKKYLD-----BEARKYLEGLKLELNKKIKD 121
Qy 84 -----LELKNLSLTPVEKNLHFIWIGQINDTAIYNVQWQVNDYTVKVFVD 132
Db 122 IIAKIEYVNTVELUKE-----IEKNYAI-----DELAN-----QSPYKVTGYIE 162
Qy 133 SNAFLINTLTKTIIVESATNTLESFRENLN-----DPEFDYN-----KPYRKRMEIYDK 182
Db 163 NKNTIYNTI-KSYPDQIYEGDIDTFYNELSSIVKEDPIDDIETKLENLSKIDNVYDK 221
Qy 183 -QKHFDIYYSK---QIEN-----PEFIDNIITKYLNSYKSLKLEALNKYIESLNKNTAN 235
Db 222 IQKMEIETVKSHLANNIETNNKLPNTILE--IKKYIYDEISKE--LNKMLEDFKKNK-EKE 275
Qy 236 NGNDIRNLEKFADEDVRLYNQELVE---RNLLAASDILRLSMLEKEDGVYLD--VDILP 291
Db 276 LSNKISDYDKREQ--LSEYKSKMLEIRNNHNSQNTVNDTKBEBAKQNYKSNHEMTTIP 333
Qy 292 GIQPDLPKSNKPSITNTSMEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSPESALS 351
Db 334 TNEDEISKIIEV-----KTMKDEILSKVNYI-----DFNKKYKETVNSEHSQFT 379
Qy 352 SKDSKSEIFLPLDDIKVSPLEVKIAFA-----NNSVINOALISLKDSY-----394
Db 380 ELTDKIK-----AEVSDKELKKCEQSFNDNKSINETKNSIEKEYQNINTLKKVDYI 432
Qy 395 --CSDL--VINQIKNRYKILNDNLNPSI-----NNSVINOALISLKDSY-----NEG 421
Db 433 KVCKSTKESITKFSKQTIKMLNQNITKVTETNSIDKSYIEKFEQILTGKQTKLENKF 492
Qy 422 TDF-----NTTMKIFSDKLASIS-NEDNMMP-----446

Db 493 TEFSLNNHANNELIKYFSDLKANLNGEENWLNQFTEKTEKFNIDKEKNIHNEIS 552
QY 447 MIKITNYLKV-GFAPDVRGTINLSPGVVTGAYQDLLMPKONSTNIHLEPELRNPF-- 503
Db 553 KIEIKIHASIVNISETEREIGINIESLNTKVFEKV--KENVTNLNKIKELKHDFSD 609
QY 504 -----PDK-----ISQTEOEITSLMSFNQARAKSQPEEYKGYFEGALGE 545
Db 610 FQKEGNIKYTKIKINDIMAVSQOIQHINGLDDI-QKSESYSVSEK-----Q 660
QY 546 DNLDPAQNTVLDDKQVSKILSKMTRNKYIHYIVQLQGDKIYSEASCNLFSDPYSS 605
Db 661 INKLEKVSNTSEISDNVE-----GIKKKQI-IVTKIDKKNIYBEINKLSB----- 707
QY 606 ILYQKNIESETAYYYVADABEIKDKYRIPIQISNKNKIKLTFIGHGKSFNTDTTAN 665
Db 708 ---ISKIEDNTS-----LEKVKOINLSYG-QNLGNLFLEQIDBEKKAE-NTIKS 753
QY 666 L-----DVSLL-----SSEIETILNLAKADISPKVIREINLLGCNMFYSIVABETYPGKLL 717
Db 754 MEAYIDDLNIIKKKSOEIEEMDI-KMDIN-KEME-----AL 788
QY 718 KIKDRVSELMPSIQDSITVSANQVEVRINEBGRKREILDHSGKWIN--KEESIIKDISK 775
Db 789 KISHDDDKKCHKSKN-----HKNISDIYDKSKIQDFRESIDINDIKNK 835
QY 776 EYISFNPKEN-----KIIVSK-YLHELSTLLOETNNAN 809
Db 836 LQKVSESQNHNSDINOCLNEVANIYILKLNKIKIKIDKVEYSETEKKNKNINDELN 895
QY 810 SSDIDLEK---KVLMECEINVASNIDROIVEGREIE-----AKNLTSDSINVKNEFK-- 860
Db 896 NSEKVIKTEGLSLKESKINSITLDDKIDECIKNVLKKNILNEETN-ITHFKN 954
QY 861 -----LIESISDLYLKHQGLDDSHFISFEDISK---TENGFIRFIN 902
Db 955 EEVNKLVSFNFNNIEMADNKSQYILEIKKNGTND-HDYNIKELKSHDKSNGYK---T 1009
QY 903 KETGNSIFITEKEIPSEYATHI-----SKBISNIKDTTFDNNVGLKVKVNLDAH 954
Db 1010 EADQNKAKIQKQKELPEQKEEVTLLNRYAYAVELKKNFKDKT-KNDSKQIIKEIK--DAH 1066
QY 955 EVNTLNSAFFIQL-----IEYNTTKESLSNLSVAMQVYAQFSTGLNTITDAS 1005
Db 1067 NCTLESSEKSKKNEIKNEKIHIEDEVANNKSKAIT-SIKSVPEPPT---KIINK 1122
QY 1006 KVELVSTALDETIDLLPTLSGLPIIATIDGVSIGAAIKELSETNDPDLRQ----- 1058
Db 1123 EIRTKSDDCLEKTNLEKQIS-NLSIDTQETKLTKENGKQLKTLBELLESKKQKNIEDQ 1181
QY 1059 -----EIEAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGISGIPSLVNNELIL 1113
Db 1182 KRELDEVNSKI-----KNIENTVN----- 1200
QY 1114 QKATKVIDYFKHIS-LAETEGAFLLDDKIIMPQDDLVLSIEDNNNSI-----TLG 1165
Db 1201 QHKKNYEIGIVEKINEIAKTNKQNESTKELIKPTIOHIS--SFNANLEGISDENUG 1258
QY 1166 KCEIMRAEGGS-----GHTLTDIDHFPSSPIYRKPMLSIYDVLNIIKKEKIDFSKD 1218
Db 1259 K---YNTENGNIYEEPIKSNLITNVLTVSKESITYNQ-----IQNKRIDTQKE 1305
QY 1219 LMVLPNAPNRVGYEMGTGFRSLDNGDKTLDRIDHYEQGYWRYPFAFADALITKL 1278
Db 1306 L--LKNIN-----VNKAKSVLDYIK--ENEFEF-----DRIVTHF 1335
QY 1279 KPRYEDTVNRINLDGNTRSFIVPVTTEQIRKNLSYFYGSG-GSYLSLSPPYNNIDLN 1337
Db 1336 KKKLNTVNDNF-----KN-EYSKVEGPDNLSINTVKNSTDBN 1374
QY 1338 LVEN-----DTWVIDVDNVKNITTESDEIOKGELIENILSKLNIEDNKIINNHTINF 1391

Db 1375 SLLNILNQTKEMYANVNVNTYYSKYBAENIFRN--IPKLANTLNK-----IKNSSGIDL 1428
QY 1392 YGDINESNRFISLTFSEILEDINII-----EIDLVSYSKYILLSGNCKMLIENSSDI 1443
Db 1429 SKDIKIA--ILSYLSDSKTEDTLFIPSPQKKTETTYTKIDSYSILLD-----ILKSOEL 1481
QY 1444 QOKIDHIGFNGEHO--KVIFYSIDNETKYNFIDYSKKEG-LFTABFSPESIIRNIYMP 1500
Db 1482 QKK-----EQOTLKLIIF-----ENRRLYEKVQATNELRGTLSDLKYKKEKILSEV--- 1526
QY 1501 DSNNLFIYSKDLKDI-----RIINKG---DVKLLIGNYFKDDMKVLSLFTIEDTNT 1549
Db 1527 ---KLLHKSNELNKLSCNFQNYDTILES SKYDQVKEKSNNYKQEKELGIDFNVTDMEE 1583
QY 1550 IKLNGVILDENGVAQILKFMNAKSAINTSNLNFLESINIKLIFYNNLDPNTEFILD 1609
Db 1584 -KFN-----NDIKVIELENNYDSSENNILOSQKQKELTNKFNKAEIKKIDDKIIEK 1636
QY 1610 NFIISSGNSIGQFELICDKDKNIQPY-----FINFKIKETSYTLVVGNRQNLIVEPSYHL 1664
Db 1637 NDLI-----DKLIETRKNCMLFTHTTLAETLKIKITDYSKFIESATFESKEFLXYI 1687
QY 1665 DDSGN-SSSTVINFSQKY-LYGDIDRYV-NKVIIAPN 1697
Db 1688 GDTNSLNDIDIATLQKYDLHQINKYVTSKLSDATN 1723

RESULT 9
G70163
Hypochemical protein BB0512 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: G70163
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White-
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: G70163
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2166 <KLE>
A:Cross-references: GB:AE001153; GB:AE000783; NID:G2688419; PIDN:AA66876.1; PID:G268842
A:Experimental source: strain B31

Query Match 4.9%; Score 424; DB 2; Length 2166;
Best Local Similarity 19.9%; Pred. No. 1.1e-07;
Matches 426; Conservative 338; Mismatches 691; Indels 684; Gaps 108;

QY 3 LVNKAQLOKQYVFRIOEDYVAILNLALEYHNMSESSVVEKYL--KLKDINLNTDYL 60
Db 8 LVNLFVLSIVLFV-YR-QYDKRSRALDKIKKFDVLTQVN-LEDFIEDTKTEINDLAVD-M 63
QY 61 NTKKSGRNKALKKEKVELTWEVLELKNLSLTPPEKNLHFIIWIGQINDTAINYNQWKD 120
Db 64 BAYORSS-IEIITKIEE--VOQTKNKNNDPFAVEKKIAY-----HDSMLKJLDEMTF 113
QY 121 VNSDYTVFVYDSNAFLINTLKTI-----VESATNNTLBSF-----RENLDNPF- 166
Db 114 KVENIQRLQVDGK--IVDKLSKTLKGFNTQIDSVESNLNSVLEKFDKANKENLESIKIA 171
QY 167 DYNKYRKRMIY--DKQKHFIIDYKSIQIEENPEFIIDNIK-----T 208
Db 172 SWEPFDTNKEKLVKIDNLNKEISLYEKDLANIEBKNDILVKGNEKLDLBFSDPLEKVE 231
QY 209 YLSNEYSKDLAL-----NKY--IEESLNKITTANNNDIRNLEKFADEDLVLNQLVE 261
Db 232 FNIGKYSKEIESSENFYENKFKLINSIELIMESVKNKINEKEDF-----ILNRLNEEQN 287
QY 262 RWNLAASADIL-----RISMLK--EDGGVYLDVDI---LPGIOPDLFKSINKPDSITNT 310

Db 1729 LKTHGLEPCTTLSPSSIVETAIDHVS-----LLQAVKTKTSRMHTLIQSLGERI 1779
Qy 1484 FTAEPNESIIRNIYMPDSNNLFYSSKDLKDIRIINKGDKVLLIGNYFKD---DMKYSL 1540
Db 1780 F-----KLFKNQSGVQSDKI-----SIDRVQLVDESDSATIDFTVIKDKORSOKSI 1826
Qy 1541 SFTIED-----TNTIKLNGVVLVDENGAQILKPM--NNAKSALNTNS 1581
Db 1827 TWGIESLAGSFRKFSASMEHVITGTGLDLD-LGMTVVSIVQVRLVEAQGGKDALAVANL 1885
Qy 1582 LAMNF---LESINIKNI 1594
Db 1886 VMNLKIALE-VSIGNV 1900

RESULT 12
Cypothetical protein PF00145c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: C71622
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: C71622
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1979 <GAR>
A:Cross-references: GB:AE001375; GB:AE001362; NID:G3845105; PIDN:AAC71819.1; PID:G384510
A:Experimental source: clone 3D7
A:Genetics:
A:Gene: PF00145c

Query Match 4.6%; Score 398; DB 2; Length 1979;
Best Local Similarity 19.7%; Pred. No. 8.3e-07;
Matches 373; Conservative 287; Mismatches 605; Indels 626; Gaps 88;
Qy 53 NNJTDNLYNTYKSGRNKALKKPKFVLTMEVLELKNSTTPVEKNLHFTWIGQINDTAI 112
Db 93 NNNDDNN 149
Qy 113 NYINQKQVNDVTVFVYVDSNA--FLINTLTKTIVESATNTNLTSPREN---LNDPEF 166
Db 150 NYESKIELEKE--LKEVDKNDNDNDYENKLEK--EDFVQKIDMLNEKENLQEKEL 205
Qy 167 DYNKPYRK---RMEIYYDKQHFIDYKSGQIENPE-----FIIDNIIKTYLS----- 211
Db 206 DINKREKKINEKEKNIKKETPHNTEKEYLEKKNKERETISIEID--IKKHLEKLIKIEI 263
Qy 212 NEYSKLEALNK-----YIBESLNKITANNNDIRNLEKFADEDLVRLYNO 257
Db 264 KKKEDLENLKLKLSKENVLEKELKGVKFKETNSLNDNIEKPKY-----KLLEY 317
Qy 258 ELVERNWLAAASDILRAISMLKEDGGVLDVLDLPGIQLPKSINKPDSITNTSWEMIKL 317
Db 318 ELEEK-----NKQIDU-----LNKQEKKEKEKEKEKE 345
Qy 318 EAIMKYKEYIPGYTSKFNDFMLDEQVRSFESALSKSDKSEIFPLPDDIKVSPLEVKIAP 377
Db 346 KEREKE-----KEYDTLIKEL-----KDEKISILEKVHSIKVEMDIEKRE 388
Qy 378 ANNSVINQALISKQVCSLDVLTQINRYKILNDNLNPSINGTDFNTMTKIFSDKLAS 437
Db 389 HNFLHMEQDLKLNKSFVKN--NNOLK-VYKCEIKNLK-----TBLKKEKELKD-IEN 438
Qy 438 ISNEDNMFMKITYNLKVGFAPDVRSTNLSGPGVYVYTGAYQDLMFKDNSTNIHLLEPE 497
Db 439 VSKEE-----INKLINQLN-----EKEKQILAFNKNHKEIHGLKEE 475
Qy 498 LRNFEPKTKISQLTQETISLWSFNQ-----ARAKSQFPEYKK 536
Db 1126-1132, 1998

Db 476 LKE-----SVKITIETOELQEMVDIKQKELDQLOEKYNAQIESISIELSKKEKEYNQYN 531
Qy 537 GYFEGALGEDNLOFAQNTVLDKDVSKILLSSMKTRNKEYIHYIVQLQGDKISYEASCN 596
Db 532 TYIE-----EINNLN-----EKUEETNKEY-----TNLQN---NTNIN 563
Qy 597 LFSKDPYSSILYQKNIEGSETAYVYVADAEIKEDKYRIPIYQISNKRNIKLTFIGHGKS 656
Db 564 MLNND-----IHLNGNIKTMTNT-----QISTLKN----- 588
Qy 657 EFNFTDFANLVDVLSSEIETILNLAADISPKYIEINLGNMPSYIYAEETYPGKLL 716
Db 589 -----DVHLINQIDK-LNNEKGLNLSKISELNV-----QI 618
Qy 717 LKIDRVSELMPSISQDSITVSANQYEV--RINEEGKEIILDHSGKWINKESIIKDISS 774
Db 619 MDLKEEKDFL-----NNQIVLSNQIDILTRMBEKENKMLEQENKYQKQEMELLRGNLKS 673
Qy 775 KEYISFNPKENKIIIVKSKYLHELSTLQEIIRNNANSSDIDLEKKVMLTECEIN-VASNID 833
Db 674 SENILNDEE-----VCDLKRKLSLASEMOMKEEHD 706
Qy 834 RQIVE-----GRIEAKNLSDSINYIKNEFKLIESDSLYDLKHQGLDDSHFISFE 887
Db 707 KKLABELKDDCDVRIREMNEKEDKINMLKBEYE-----DKINTLKEQN-----E 750
Qy 888 DISKTENGFRIRFINKETGNSIFITEKEIFSEYATHISKEISNIDKTFIDNVNGLVKVK 947
Db 751 D-----KINTLKEQ--NEDKINTLKE-----EY-----EHKINTMKEEYEHKIN-----T 788
Qy 948 VNLDAAHEVNTLNSAFFTQSLIEYNTTKESLSNLSVAM-----KVQVYAQLFSTGLNTI 1001
Db 789 LNEQNEHKINTLINE-----QNEHKINTMKEEYEDKNTLNEQNEQDNLSKBEYENKINQI 844
Qy 1002 TDAS--KVVELVSTALDETIDLLPFLSEGLPIIATIIDGVSLGAAIKELSETNDPPLLQOE 1059
Db 845 NSNNEIKIKDVVNEIIEVDKLVKTLDEKKQFQKDEINVAHIKAHEKE----- 892
Qy 1060 IEAKIGIMAVNLTAASTAIVTSALGIASGFSITLLVPLAGISAGISPLVNNELLQDKATK 1119
Db 893 -----QILLTEM-----BELKCO-RDNK 909
Qy 1120 VID-YFKHISLAETGAFITLDDKIIMPQDDLVLSEIDFNNNNSITLGCETWRAE--GGS 1176
Db 910 YSDLEYKYIKLAKSICMIINECCDDIENEDIIRIEEYINNNGKLGK-EVEKEKGRHS 968
Qy 1177 GHTLTDDIDHFPSSPSITYRKPWLSIYD--VLNKKKEKIDFSDKOLMVLNPNRNVFGE 1233
Db 969 SPNILKSKEKFPKN-----SIEDKSHELKKKHEKDLKSKDKET--EEKNK----- 1011
Qy 1234 MGWTFGFSLNDGDKLADRIR-----DHVEGQFYWRYPFAFIADALITKLKPRY 1282
Db 1012 -----KIKELNNDIKKLQDEILVYKQNAQOQVDHKKKSWI-----LKQKSKKEKI 1057
Qy 1283 EDTNVRIINDGNTRFSI-----VPVITTEQIR-KNLSYFYGSGGYSLSLSPYNNMID 1335
Db 1058 KDKENQINVERNEEKDLKKDKODEIRILNEELVKYKTI-----LYNLKDKPLLQND 1108
Qy 1336 -NLVENDTWJD-----VDNVVNNTIETSEIQKG-----ELIENILS-----KLNTED 1379
Db 1109 LLSKIDINSLATINEGCMVDKIEHILYDEDEINKRSRNLFLQKNIEICSLTTEVMEINLNNK 1168
Qy 1380 NKIILLNHTINFG-----DINESNRFTSLTFSILEDNIIIE-----I 1418
Db 1169 NELIBENKNLNVDOGKKKLKQVKEKKEKLNKQUTCKNQIDELNEEVEKLNENI 1228
Qy 1419 DLVSKS-----YKILLSGNCMKLIENSDDIOQ---KIDH----- 1449
Db 1229 ELITTYNDLNNKFKDMKENNLMKLEDNEDNKKMSKIDDMKEIKYREDEKKNLNEIN 1288
Qy 1450 -----IGFNGEHOKY-----IFVSYDNTETKNGF-IDYSKK----- 1480
Db 1289 NLKKKNEDMCIKYNEMNIKYGDICVKYBEMSLTYKETSLLKQBQIKVKYDEKCSQYDEIRF 1348

Qy	1481	---	EGLTAFBSNE--SIIRNIYMPDSNNULFIYSSKDLKDRIINKGDVKKLLIGNYFK--	1533
Db	1349	QYDEKCFQYDEINKKYGALLNI---	NITNKWDSKVDRNNNEIISVDNKVEGIANYLKQI	1405
Qy	1534	-----	DMKVSLSFT--IEDTNTIKLVGVLDENGVAQILKFM-----NNAKS	1574
Db	1406	FELNEEIIIRLKEINKISLAYSNEBKNSYDINMKHIOEQ----	LLPLEKTNKENEK	1460
Qy	1575	ALANTSLSMN-----	PLESINI-KNIFYNNLDPN---IEFILDTNFIISG	1615
Db	1461	IINITSQYDAYKKKSDSKGQAQVDDVNIYGNISNNIRNTEYKYEEFMDTNIIEKN	1520	
Qy	1616	SNISQPELICDKDK--NIOPYFNFKIKET-----	SVTLVGNRONLIVE---PSYH	1663
Db	1521	GMHLSKYIHLEENKFCMKIIVENENIKSNKIIGLYNYSRYGLREDLCKEIVPS--	1578	
Qy	1664	LDPSGNTSSTVINPFSQKYLXIDRYNVKVI	1694	
Db	1579	--KIGNISNKNENNKKN--NTCDGYDEKVTI	1606	

RESULT 13
 T18427
 hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18427
 R:Lawson, D.; Bowman, S.; Barrell, B.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18427
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-3724 <LAW>
 A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1
 C:Genetics:
 A:Introns: 307/1; 1545/2
 A:Note: C0335C

Db	936	EITNKELDPLEINTQNEFIENLDIKKKYKYNDFNDADKMFYEMNKILN-----KOM	998
Qy	463	RSTINLSGPGVYTGAUYODLLMFKDNSTNIHLLPELRNFPFKTKISQLTQEQBITSLWSF	522
Db	989	KKN-----KEQFFTKTDTFGLSQSHKKYKNKGEEKDKDKNNEEKNYLYDE	1035
Qy	523	NQARAKSQFBEYKKGYPGEGALGED-DNLDPAQNTVLDKQYV-----SKKIL-----SSM	570
Db	1036	NQV-----YSVLYSDHKIEQDIOIHISIQTNICDENNIEQINEENSKKGVRSIGTDM	1087
Qy	571	KTRNKVEYHIVLOLGD---KISYEASCNLFSDPYPSSILYQKNIEGSETAYYYYVADAE	627
Db	1088	ENKNDMENKDMKENDMEKNDIEKKNDEKKND---MEKKN-----DME	1130
Qy	628	IK-EIDKYRIPYOISNKRNIKLTFIGHGSEFNFTDPAULDVDSLSEIETILN-----	680
Db	1131	KNDMEKKN---DMENENNME-----NKSIDIENENKMEYKONMKENKMIENNTIENN	1181
Qy	681	LAKADISPKYIEINLGCNMFYSIYAETYPGKLL---LKIKDRVSELMPISODSIT	736
Db	1182	MIENMIENMIENMIENMIENMIENMIENMIENMIENNTKIAYMNE-----NKNNIK	1236
Qy	737	V---SANOYEVRINEEGREILDHSGKWINKEESIHKD-----ISSKEYISFNP	782
Db	1237	IEDDAKNQINQVNNIQNVKHVKKQNDLINNV-DIINDVLKSDDFENMINSKE---INI	1292
Qy	783	KENKIIVKSK---YLHELSTLLQETRNNANSDDILEKKVMLTECEINVASNIDRIVE	838
Db	1293	KDFKCVKDNNDICLUNELNK---QSNYSNTEKCIDIH-----DFDYVENKO---VH	1339
Qy	839	GRIEAAKNLTSDSINYIKNEFKLIESISDLYDLKHQNGLDGSHFISFEDISKTENGFR	898
Db	1340	DKIHEDRKEYCD-----ESKLKYLPADDIKKWR-----SPKINKKSKRENFLII	1384
Qy	899	RFINKETG---NSIFTE-TEKEIPSEYATHISKE-IGNIKDTTFDNV-NGKLV-KKNULD	951
Db	1385	SYNNEEYIFLYIKKLKVEIIEPSIILNKNMHIEKQFLSPGLKYINDNLKEGKCITSSKI	1444
Qy	952	AAHE-VNTLNSA--FFIOSLIENTTK-----ESLSNLSVAMKVQVYAQLFSTGLNTIT	1002
Db	1445	LMLEYMNTISKNGVYLYPNYNNIIOHLDYFLHUKLGUSTIKYINVHPLNFINFLKVN	1504
Qy	1003	DASKVVELVSTALDETIDLPLTLLSEGI-PIIATII---DQVSLGAA-----IKELS	1049
Db	1505	D-QKDVNIYDLLDDQKIIVSNPFKMGPKFCLILKKNHIVSIHSGKHFNPKQIYRLLS	1563
Qy	1050	ETNDPLL-----ROEIAKIGIMAVNLTAASTAIVTSALGASGSFILLVPLAGISAGI	1103
Db	1564	KFNDKQIDSSNNYKVEEVEEK-----	1583
Qy	1104	PSLVNNELIQQDKATKI-----DYFKHISLAETEGAFTL-----	1138
Db	1584	---HSELFAMDKNEEKYKVKENQENBHDHKSVCSEIYLSIS---SEKINTYDGEDEQD	1636
Qy	1139	-----	1138
Db	1637	EQDEQDEQDQDEBEDEQEEDEQEEDEDEDEDEDEDEDEDEDDDDNDNDDDE	1696
Qy	1139	-----LDDKIIMPQDLVLVS-EIDFNNSI-TLGKCEIWRAGGSGHT-----	1179
Db	1697	DDNDNDDDDDEYDDHDKPSSDATQNISINIKSIIEBINNYENTNKNKDKHKTDES	1756
Qy	1180	--LTDIDIDHFF--SSSITTYKRWLSIYDVNLTK--EKIDFSKOLMWL-----	1222
Db	1757	NLNMKEINFYKCTSPDKYYK-----NYYN--NLUKYCDMEHEKDSMLIKHLSVLLKDE	1810
Qy	1223	PNAPNRVFCYMGWTPGFRSLONGTKLLDR-----IRDHYEQFYRYFAFI	1270
Db	1811	KENIKNNEGEB---SIN---ILIDKNMKSDQEVENEKOHNLNIEY-----	1852
Qy	1271	ADALITLKPREDTNNRINLDGNTRSFIVPVITTTBQIRKNSLYSFIYGGSGSYLSLSPY	1330

Db 1853 IENLSHLRHKYDIEMIPIDIVGNIR--LYKLFYINMIKNN--YSYLKQNNFF-----PY 1903
Qy 1331 NNMIDNLVENDTWIDV-----DNVKNITIESDEIQGELJENILSKLNIENDKIL 1384
Db 1904 KYGKRGKAKGMLMINIKRKHEDYDMLKK--KNGNEKTG--NNKKKKNNKEDNDYI 1957
Qy 1385 NNHTINFYGDINESRFLSILFSLIEDINIIIEIDLVSYSKYL---LSGNCMKLIENS 1441
Db 1958 NNNNNNNNNNDNN-----INV-----GYKRMQNNINPNLYEIKNN 1996
Qy 1442 DIQOKIDHIFNGEHO-----KYIPYSYIDNETKNGYFIDYSKKEGLPTAFSNEISIR 1495
Db 1997 -----FNLYHRYSLFYKYPVSIND-----NKIEGKK--PCVYKKN-----2035
Qy 1496 NIYMPDSNNLFI-----YSKOLKDI-----R.IKNGDVKLL-----1527
Db 2036 -IYMLNNHDMYIEFLKHPKPYASKVKESVCSLVIYFPNMDIQKKIYLLTYEKLHI 2094
Qy 1528 -IGNYF-----KDDMKVS-----LSFTIEDTNTIKLN-----1553
Db 2095 DIGKYPKQYFFLSHQDKEQKNNKSNNDVDDNKIEDAHIFEHRIIEGKTIISPLFIV 2154
Qy 1554 -----GVYLDENGAQILKFMNNAK-SALNTSLSLMMFLE-----1587
Db 2155 RVICIHMKKIHOPIYIIEIITSIKLKHNFLLIMKKKILNQCEIKKKMKKKKK 2214
Qy 1588 -----SINKNIFYNNLDNIEF-----ILDTFIISGNSISIQFELICDKDKNIQPYFIN 1638
Db 2215 KKKKKKIPEKKLVNNEKKNLFLFSSIMKIKFL-----YNLIITIKQKNK--VS 2263
Qy 1639 FKIKETSITLYVGNRQNLIVPSYHL-----DQSGNISSTVFNFSQKLYGIDRYNKKV 1692
Db 2264 FK-----NVLDKLSHTINIKDAKIDNLLVLEDMFPEKKKYLNVK 2306

RESULT 14
T18501
hypothetical protein C0760c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18501
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18501
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-3394 <LAW>
A;Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331922; PIDN:CAB11140.1
C;Genetics:
A;Map position: 3
A;Note: C0760c

Query Match 4.5%; Score 386.5; DB 2; Length 3394;
Best Local Similarity 19.0%; Pred. No. 4.4e-06;
Matches 401; Conservative 319; Mismatches 699; Indels 687; Gaps 99;
Qy 6 KAOLQKVVYKFRIODEYVAIINALEEVH---NMSESV-----VEKYLKLDINNLTLD- 57
Db 699 KIYLEKLLNEKLYIKD-----IEKKLDLHGVIKKNKEDIYILQVEKQTLIKVISSVYDY 754
Qy 58 -----NYL-----NTYKSGRNKALKFKLEYLTWVELELKNLSLTPVEKNLHFIMIGGOI 107
Db 755 TKMESENHIFKQMTTWKMLNNVHMSNKKDYNNQNNQNNQNNQNNQNNQNNQNNQNNQNI 814
Qy 108 NDTAINYNQWKD-----VNSDYTVKVFYVDSNAFLINTLTKTIV---146
Db 815 ENNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNN 874
Qy 147 -----ESATNNTLESFRENLDNPDYDNKFKYRKMEIYDKQ 183
Db 875 KDEKKTSNIKESISLKKKHTKSTINNNDN--NNEDDDNDMLSVMSYNDVDVKNKR 932

Qy 184 KHFIDYKSOIEENPEFIIDNIIKTYLSNEYSKOLEALNKYIEBSLNKITTANNNGDIRNL 243
Db 933 K-----ENKKEILIEHVSPSSFSN-----NEYIAHSFNSILLQLSNYIFNI 973
Qy 244 EKFADEDLVRLYNOELVERMNLAAASDILRISM-----LKEDGGVYLDVDILPGIQPDLF 298
Db 974 E-----CKOMEYFKNSLLSYVDYTTITIELFVRLKKNYINIFSIQILGTQYPSIL 1024
Qy 299 KSINKPDSIITWTSWEMIKLEAIMKVKYIPCYTSKFNDFMLDEEVQVSFESALSXSDD-- 355
Db 1025 OKLHDGIYCLKDNKKNNKNNQDNGKQEDDDGNNKKKNDG--DNKSQEDDGNKKNNQDGN 1083
Qy 356 KSEIFLPDDIDKVSPLVKIAFANNVINOALISLKDSYCSDDLVIINQIKNRYKIKLNDNLN 415
Db 1084 KSQ-----EDD-----YGNKKKND--DDDDSYKIELIVDELNKKCKNYTDEEL 1126
Qy 416 PSINEGDTFNTMK---IPSDKLASINEDNNMFIKITN---YLVK-----GP 458
Db 1127 YELMKGSDFDIFKKYKQNFYLNHFININNIPSTIISFHHNIEDKYKYVRYFNLFNFNF 1186
Qy 459 APDVRSTINLSGPV-----YTGAYQDLL-----MFKDNSTN-----IHLPE 497
Db 1187 -----SNVELSFDLIRRFDKILRLTKYEQLEENVEKIKNKEKEEYLHACIKELMN 1241
Qy 498 LRNFEPFKTKISQLTQEITSLWSFNQAR--AKSQFEYK-KGYFEGALGEDDNDL--550
Db 1242 LERYNNEKIIIDDEQINEKEKINIINEKYLILEXEYEQKNIFINA--QIENLEKEKK 1299
Qy 551 FAQNTVLDKDYVSKKI-----LSSMTNRNKEYIHYIVQLOQDKLISYBASNLSFKDPYS 604
Db 1300 QLQEBIIQKDMINVKLNEKNDIIKIYEKQYLLHTLLOENKDSHNY-----LKDKFE 1352
Qy 605 SILQKNEGSETAYVYVADAETKIDKYPQOISNKRNIKLTFFIGHGKSEFNTDTFA 664
Db 1353 NLL---NL--NEKLKYDH--DISLNKINTLWLEKNNKK--TF-----HMNLRVE 1396
Qy 665 NLDVDSLSSEIETIINLAKADISPKYIEINLLGCMFYSYI-----YAE 708
Db 1397 NNNLLKMKELQNKYNIKKELNERIKQINVRNNVSTLSLRDNSTRGRSHQINNMWYN 1456
Qy 709 ETPYQKL--LLKIKORVSEL-----MPSIQDSITVSANQYEVRIEE--749
Db 1457 NTHLFGMGASKINNINISNLYSNMTHSHRGSIIKNKEDASGNSIQARMNNKSDTNIIN 1516
Qy 750 -----GKREILDHSGKWINKEE 766
Db 1517 NIHTNDINNMMNNNNNTLNSINSHLYPPFPFNNVNSPKVMGCMCDVTLASG--VNKK 1574
Qy 767 SIIDISSKYSIFNPKENKIIVKSKYLHELSTLLQEIIRNNANSDDIDL--EKKVMLTEC 824
Db 1575 DFLNLNEENESF--LEVEIRKS-----LQBELCDKSEELIKGKKNLITCI 1623
Qy 825 EI-----NVAENIDR--QIVEGRIEEAKNL-----TSDSINYI-----855
Db 1624 ETWKCFCKNSKEEISRLKEICEQLEKHEKFLINKSNEDKLYINSLSLCDKDKYDIWV 1683
Qy 856 -----KNEF-KLIESISDSLYD--LKHONGLDDSHFISFEDISKTENGRIFRINKE 904
Db 1684 KDIKNNMNEIDKNNNDINEKSEYELKLKHN-----NNLINEMNILK-----NKE 1729
Qy 905 TGNSTIFETEKEIFEYATHISKEISNKTDTIFONVNGKLVKVNLDAAHVAENTLNSAFF 964
Db 1730 TENNNIKQKEE---DYIKLIKOKNTIQNEVD-----LLEKYN-----EYVVKNNMLY 1775
Qy 965 IOSLIEYNTTKESL-----SNLSVAMKVQVYA-QLFSTGLNTITDASKVVELVSTALDETI 1019
Db 1776 NDMVLLKHEKEEIEPLKENIKILOKONTYLDNMFKNQINYVDN-----NLLKNRLDOLF 1830
Qy 1020 DLLPTLSGLPIIATIIDGVSIGRAIKELSETNDPLLRQ-----EIEAKTIGIMAVNLTA 1073
Db 1831 NNQLOQLKHL-----DINQKHLQOLKYDYIEIKERLKEKTINK 1870
Qy 1074 ASTAIVTSALGIASGFSILLVPLAGISAGIFSLVNNELILODKATKVIDYFKHSLAETE 1133

Db 922 NTSEHEKDER-----DGKYLEASINDYMSDDKKKRYVDSIESLRGSKI-- 965
Qy 864 SISDSLYDLKHQNL-----DDSHFISPEDISKTEGPRIRPINKETGNSIETEK 915
Db 966 -KNDQIYOGGHSSLLYYDDNNDDNNNNYSSSSNNHYIYLTNDKELNNDNFNNNL 1024
Qy 916 EIFSEYATHISKE---ISNIKOT-----IFONVNGKLVKYNLDAAEHVNTLNSAFFIQ 966
Db 1025 EINNSQNKVIEKNLEYINNKKLTGTSNYEQSNNTNSKDEHNISSSDKSKEDTLNLS--RK 1082
Qy 967 SLIEYNT-TKESLSNLSVAMKVQVVAQLFSTGLNITTDASKVVELVSTALDETIIDLPTL 1025
Db 1083 SSVEYNNKILQSTNSKSLNGAVE--NNLFS-----GKKKNGKGTVLKD-IEHINDI 1130
Qy 1026 SEGLPITATIIOGVSLGAIAKELSETNDP-LLRQETEAIGIMAVNLTAASTAIVTSALG 1084
Db 1131 QOKYP-----EDLNINCVNKVVIENEERKHLPLELE-----YNLVSDEKF----- 1171
Qy 1085 IASGFSILVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGFTLLDDK-- 1142
Db 1172 -----GLNKIKNNNII-----YMKH-----QNYHNLYDDNQK 1199
Qy 1143 --IIMPQDDLVLSEIDFNNSITLGKEIWRAGSGGHTLTDDIDHFFSPSPSITYRKPLW 1200
Db 1200 KHLIPDTNKNV--SIQRNNNSIVIKTNHVEVEKNNKQORNYD-----NFTCDKXKK 1249
Qy 1201 SYDVLNIIKKEIDFSKDLMLVLPNAPRVFGYEMGWT-----PGFRSLDNDGK-LIDRIR 1255
Db 1250 IYNNIIN--SDKDIYHNNIYYTNEKEGIGNIHLNRNDKIDITNFELLKLDGVKEFLDTFK 1307
Qy 1256 DRYEGQFYWRYPAFIADALITKLKPRYEDTNVRINLDGNTRSFIVPVIITEQIRKNLSYS 1315
Db 1308 DSY-----IDCHNKKENI----- 1320
Qy 1316 FYGSGGSYSLSPYNNIDNLNLVENDTWVIDVNVVKN-----ITTESDEIQGELIENI 1371
Db 1321 -----LNMTNKN-----KEDHQIIDVADKI FNETNMTWNNKIYDDK----- 1358
Qy 1372 LSKLNTEDNKIILNNTINPYGDI-----NESNRFISLTFSILEDINIIEIDLVSXS 1424
Db 1359 ---NVHEKK-CTHNDVIHNMMDILSTSIKNNEENLF-----IDTYQKQ 1397
Qy 1425 YKI-LLSGNCMKLIENSIIQKIDHIGFNGEHQKVIYFYSIDNETKYNG----- 1473
Db 1398 NRIGDIYMRINILQEDDDDDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1457
Qy 1474 FIDYSKKEGLFTAESNES---IIRNIYMPDSNNLFIYSSKDKIDIRIINKGDKVLLIGN 1530
Db 1458 FSDFIEKKKIKIKKNKESYHKIDESLLSNEKN-----KVSILLIN 1499
Qy 1531 YPKDDMKVLSFTIETDTNVIKLVGVYLDENGVAQILKFMNNAKSALNTSNLSMNFLESIN 1590
Db 1500 NKKDSSVD-----NNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1549
Qy 1591 IKNIIFYNNLDPNTEFTIDTNFI-----ISGSNISIQFELICDKDKNIQPYFINFKIKE 1643
Db 1550 DNN-----NNNNNDSFSKDNLLNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1606
Qy 1644 TSYTLVGNRONL-----IVEP-----SYHLLDSGNSISSTVINFSQKLYGIDR 1687
Db 1607 VSVSPLINHPNLSMRKORTIEFLKINGKNKLIKDLAKIKQOVERKIRKYIQMDQ 1664

Search completed: November 5, 2003, 19:32:50
Job time : 42.125 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:27:13 ; Search time 13.875 Seconds
(without alignments)
7069.698 Million cell updates/sec

Title: US-09-126-816B-6_COPY_1_1020
Perfect score: 5192
Sequence: 1 MNLVNAQLQKVVYKFRIQ.....ITDASKVVELVSTALDETID 1020

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5171	99.6	2364	2 I40884	cytotoxin L - Clos
2	4086	78.7	2366	2 S10317	toxin B - Clostrid
3	4073.5	78.5	2367	2 S70172	toxin B - Clostrid
4	2783	53.6	2710	2 A37052	toxin A - Clostrid
5	1424.5	27.4	2178	2 S55805	alpha-toxin - Clos
6	408.5	7.9	3169	2 T00296	toxin B - Escheric
7	362	7.0	3225	2 D81702	adherence factor T
8	350	6.7	2401	2 T28676	rhostry protein -
9	345.5	6.7	1127	2 T28317	ORF MSV156 hypothe
10	330	6.4	3724	2 T18427	hypothetical prote
11	318.5	6.1	2166	2 G70163	hypothetical prote
12	316.5	6.1	3255	2 G81702	adherence factor T
13	315.5	6.1	3394	2 T18501	hypothetical prote
14	311	6.0	1302	1 JC6009	surface-located me
15	310	6.0	2269	2 T28677	rhostry protein -
16	310	6.0	3335	2 H81702	adherence factor T
17	302	5.8	2819	2 A90551	conserved hypothet
18	293.5	5.7	1252	2 B42771	reticulocyte-bind
19	288.5	5.6	1447	2 F82909	hypothetical prote
20	286.5	5.5	1979	2 T16222	hypothetical prote
21	285.5	5.5	4688	2 F82885	hypothetical prote
22	285	5.5	1306	2 T28313	ORF MSV152 probabl
23	280	5.4	1516	2 E71619	RAD2 endonuclease
24	275	5.3	1939	2 T18372	repeat organellar
25	273.5	5.3	2829	2 A42771	reticulocyte-bind
26	272	5.2	980	2 E71606	hypothetical prote
27	271	5.2	1711	2 T18429	hypothetical prote
28	266.5	5.1	1121	2 F71613	hypothetical prote
29	261.5	5.0	1183	2 F90559	conserved hypothet

30	261	5.0	1017	2 D90550	vsaa-like (mycopla
31	258.5	5.0	1712	2 C71618	hypothetical prote
32	257.5	5.0	1162	2 A47708	progenitor toxin n
33	257	4.9	963	2 C90535	conserved hypothet
34	256.5	4.9	1875	2 S38173	myosin-like protei
35	256	4.9	1090	2 H86806	hypothetical prote
36	255	4.9	1365	2 T30822	limpi protein - Myc
37	255	4.9	1817	2 D71606	hypothetical prote
38	255	4.9	2485	1 H71621	serine/threonine-s
39	254	4.9	2136	2 A05037	hypothetical prote
40	250	4.8	1640	2 A24594	probable major sur
41	249	4.8	1639	2 S05603	major merozoite su
42	248.5	4.8	1631	1 SAZQK1	multiple banded an
43	247.5	4.8	1103	2 H82884	hypothetical prote
44	247	4.8	1679	2 S48385	hypothetical prote
45	246	4.7	1619	2 T18499	hypothetical prote

ALIGNMENTS

RESULT 1

I40884
cytotoxin L - Clostridium sordellii
C:Species: Clostridium sordellii
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40884
R:Green, G.A.; Schue, V.; Monteil, H.
Gene 161, 57-61, 1995
A>Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium s
A:Reference number: I40884; MUID:95369733; PMID:7642137
A:Accession: I40884
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2364 <RES>
A:Cross-references: EMBL:X82638; NID:gi1000694; PIDN:CAA57959.1; PID:gi1000695
C:Superfamily: cpl repeat homology
C:Keywords: cytotoxin

Query Match 99.6%; Score 5171; DB 2; Length 2364;
Best Local Similarity 99.7%; Pred. No. 9.2e-187;
Matches 1017; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MNLVNAQLQKVVYKFRIQEYVAIINALBEYHNMSSESVVEKYKLKDNNLTNDYL	60
Db	1	MNLVNAQLQKVVYKFRIQEYVAIINALBEYHNMSSESVVEKYKLKDNNLTNDYL	60
Qy	61	NTYKSGRNKALKKPKKYLTMVEVLELKNNSLTPPEKNLHFIWIGQINDTAINYNQWKD	120
Db	61	NTYKSGRNKALKKPKKYLTMVEVLELKNNSLTPPEKNLHFIWIGQINDTAINYNQWKD	120
Qy	121	VNSDYTVKVFYDSNAPLNTLTKTIVESATNTLTESFRENLDNPEFDYKFKRMEIY	180
Db	121	VNSDYTVKVFYDSNAPLNTLTKTIVESATNTLTESFRENLDNPEFDYKFKRMEIY	180
Qy	181	DKQKHFIDYKSOIBENPEFIIDNIKTVLSNEYSKDLKALNKYIEESLNKLTANNNDI	240
Db	181	DKQKHFIDYKSOIBENPEFIIDNIKTVLSNEYSKDLKALNKYIEESLNKLTANNNDI	240
Qy	241	RNLKFADEDLVRLNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDPLFKS	300
Db	241	RNLKFADEDLVRLNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDPLFKS	300
Qy	301	INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQRFESALSOKSEIF	360
Db	301	INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQRFESALSOKSEIF	360
Qy	361	LPDDIKVSPLEVKTAFAFANNVINQALISLKDYSDDLVINQIKRYKILNQNLPISNE	420
Db	361	LPDDIKVSPLEVKTAFAFANNVINQALISLKDYSDDLVINQIKRYKILNQNLPISNE	420
Qy	421	GTDFNTTWKIFSDKLASINEDNMFMKITNYLVKGFAPDVRSTINLSGPGVYTGAYOD	480

Db 421 GTDFNTMTKIFSDKLASIGNEDNMFMKIITNYLVKGFAPDVRSTINLSGPGVYTGAYQD 480
Qy 481 LLMFKDSTNTHLLEPELNFEPKTKISQLEQETISLWSNQARAKSQFEYKKGYPE 540
Db 481 LLMFKDSTNTHLLEPELNFEPKTKISQLEQETISLWSNQARAKSQFEYKKGYPE 540
Qy 541 GALGEDNDLFAQNTVLDKDYVSKILSSMKTRNKEYIHYIYVQLOQDKISYEASCNLFPSK 600
Db 541 GALGEDNDLFAQNTVLDKDYVSKILSSMKTRNKEYIHYIYVQLOQDKISYEASCNLFPSK 600
Qy 601 DPYSSILYQKNTGSETAYYYVADAEIKEDKIRIPYQISNKRNIKLTFIGHGKSEFNT 660
Db 601 DPYSSILYQKNTGSETAYYYVADAEIKEDKIRIPYQISNKRNIKLTFIGHGKSEFNT 660
Qy 661 DTFANLDVDSLSEIETILNLAKADISPKYIEINLGCNMFYSIYARETYPGKLLKTK 720
Db 661 DTFANLDVDSLSEIETILNLAKADISPKYIEINLGCNMFYSIYARETYPGKLLKTK 720
Qy 721 DRVSELMPSISQDSITVSANQYEVRIINEREGKREILDHSGKWKNEKESIIKDISKEYISF 780
Db 721 DRVSELMPSISQDSITVSANQYEVRIINEREGKREILDHSGKWKNEKESIIKDISKEYISF 780
Qy 781 NPENKIIIVKSKYLHELSTLLQBIIRNANSSDIDLEKKVMTCEINVASNIDRQIVEGR 840
Db 781 NPENKIIIVKSKYLHELSTLLQBIIRNANSSDIDLEKKVMTCEINVASNIDRQIVEGR 840
Qy 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHFIISFEDISKTENGPRIF 900
Db 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHFIISFEDISKTENGPRIF 900
Qy 901 INKETGNSFIETEKIEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEHVNTLN 960
Db 901 INKETGNSFIETEKIEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEHVNTLN 960
Qy 961 SAFFIQSLIEYNTTKESLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 SAFFIQSLIEYNTTKESLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 2
S10317
toxin B - Clostridium difficile
C:Species: Clostridium difficile
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S10317; S21894; S22434
R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.
Nucleic Acids Res. 18, 4004, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.
A:Reference number: S10317; PMID:90326540; PMID:2374729
A:Accession: S10317
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-2366 <BAR>
A:Cross-references: EMBL:X53138; NID:g40442; PIDN:CAA3298.1; PID:g40443
R:Richel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
submitted to the EMBL data Library, July 1991
A:Description: Comparative analysis of Clostridium difficile toxins A and B.
A:Reference number: S21894
A:Accession: S21894
A:Molecule type: DNA
A:Residues: 1271-2366 <BIC>
A:Cross-references: EMBL:X60984; NID:g40445; PIDN:CAA43299.1; PID:g40446
R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
Mol. Gen. Genet. 233, 260-268, 1992
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.
A:Reference number: S22434; PMID:92293124; PMID:1603068
A:Accession: S22434
A:Molecule type: DNA
A:Residues: 1791-2366 <VON>
A:Cross-references: EMBL:X60984
C:Genetic:
A:Gene: toxB
C:Superfamily: cpl repeat homology

C:Keywords: cytotoxin
Query Match 78.7%; Score 4086; DB 2; Length 2366;
Best Local Similarity 77.5%; Pred. No. 5.4e-14;
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;
Qy 1 MNLVNAQLOKQVYVVKFRIQDEYVAIILNAAEEYHNMSSESVVEKYKLKIDINNLTDNYL 60
Db 1 MSLVNRKQLEKXANVFRFTQDEYVAIILDALEEYHNMSENTVVEKYKLKIDINSLTDIYI 60
Qy 61 NTKYKSGRNKALKKPEYLTMEVLKKNLSLTPVEKNLHFIWIGQINDTAINYNQWKD 120
Db 61 DTKYKSGRNKALKKPEYLTMEVLKKNLSLTPVEKNLHFIWIGQINDTAINYNQWKD 120
Qy 121 VNSDYTVKVFVDSNAFLNTLTKTIVESATNTLTESFRENLDNDFYKPKRKEIY 180
Db 121 VNSDYTVKVFVDSNAFLNTLTKTIVESATNTLTESFRENLDNDFYKPKRKEIY 180
Qy 181 DKQKHFIIDYKSOIBENPEFIIDNIITKYLSEYKDEALNKYIEESLNKITANNNDI 240
Db 181 DKQKFNINNYKAQRENPELLIDDIIVKYLSEYKDEALNKYIEESLNKITANNNDI 240
Qy 241 RNLEKFADEDLVRLYNQBELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFS 300
Db 241 RNFEFEKNGESPNLYEQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFS 300
Qy 301 INKPOSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVORFESALSSKSKSEIF 360
Db 301 IEKPSVTVDFWMTKLEAIMKYKEYIPGYTSKNFMDLDEEVORFESALSSKSKSEIF 360
Qy 361 LPLDDIKVSPLEKVIKAFANNSVINQALISLSDSYCSDLVINQIKNYKILNDNLNPSINE 420
Db 361 SLSGDMELASPLEKVIKAFANNSVINQALISLSDSYCSDLVINQIKNYKILNDNLNPSINE 420
Qy 421 GTDFNTMTKIFSDKLASIGNEDNMFMKIITNYLVKGFAPDVRSTINLSGPGVYTGAYQD 480
Db 421 DNFNTTNTTFTDSTMAEADNGRFMMELGKYLVRGFPDVKTTINTLSGPEAYAAAYQD 480
Qy 481 LLMFKDSTNTHLLEPELNFEPKTKISQLEQETISLWSNQARAKSQFEYKKGYPE 540
Db 481 LLMFKDSTNTHLLEPELNFEPKTKISQLEQETISLWSNQARAKSQFEYKKGYPE 540
Qy 541 GALGEDNDLFAQNTVLDKDYVSKILSSMKTRNKEYIHYIYVQLOQDKISYEASCNLFPSK 600
Db 541 GALGEDNDLFAQNTVLDKDYVSKILSSMKTRNKEYIHYIYVQLOQDKISYEASCNLFPSK 600
Qy 601 DPYSSILYQKNTGSETAYYYVADAEIKEDKIRIPYQISNKRNIKLTFIGHGKSEFNT 660
Db 601 DPYSSILYQKNTGSETAYYYVADAEIKEDKIRIPYQISNKRNIKLTFIGHGKSEFNT 660
Qy 661 DTFANLDVDSLSEIETILNLAKADISPKYIEINLGCNMFYSIYARETYPGKLLKTK 720
Db 661 DTFANLDVDSLSEIETILNLAKADISPKYIEINLGCNMFYSIYARETYPGKLLKTK 720
Qy 721 DRVSELMPSISQDSITVSANQYEVRIINEREGKREILDHSGKWKNEKESIIKDISKEYISF 780
Db 721 DXISELMPSISQDSITVSANQYEVRIINEREGKREILDHSGKWKNEKESIIKDISKEYISF 780
Qy 781 NPENKIIIVKSKYLHELSTLLQBIIRNANSSDIDLEKKVMTCEINVASNIDRQIVEGR 840
Db 781 NPENKIIIVKSKYLHELSTLLQBIIRNANSSDIDLEKKVMTCEINVASNIDRQIVEGR 840
Qy 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHFIISFEDISKTENGPRIF 900
Db 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHFIISFEDISKTENGPRIF 900
Qy 901 INKETGNSFIETEKIEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEHVNTLN 960
Db 901 INKETGNSFIETEKIEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEHVNTLN 960
Qy 961 SAFFIQSLIEYNTTKESLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 AAFPIQSLIEYNTTKESLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

Query Match	78.5%	Score 4073.5	DB 2	Length 2367
Best Local Similarity	76.9%	Pred. No. 1.6e-145		
Matches	785	Conservative 113	Mismatches 122	Indels 1
				Gaps 1
Qy	1	MNLVNKAQLOKVVYKPERIODEVAITLNALEEYHNHNSSESVVEKYLLKADINNLTDNYL	60	
Db	1	MSLVNRKQLEKMANVRFOVEDEVAILDALEESYHNHNSENTVVEKYLDKADINSLTDIYI	60	
Qy	61	NTYKKSGRNKALKKFKEYLTMVELEKNSNLTTPVEKULHFIIWIGGOINDTAIYINQWKD	120	
Db	61	DTYKKSGRNKALKKFKEYLVEILEKNSNLTTPVEKULHFIIWIGGOINDTAIYINQWKD	120	
Qy	121	VNSDYTVKVVYDSNAFLINTLKTIVESATNWTLESFRENLDNPEFDYNNKPYKRMETIY	180	
Db	121	VNSDYNVNVFYDSNAFLINTLKTIIIESASNDTLESFRENLDNPEFNHTAFFRKMQLIY	180	
Qy	181	DKQGHFDIYKYSQIEENPEFIIONIIKTYLSNEYSKDLEALNKYIEESLNKIITANNNDI	240	
Db	181	DKQONFINYYKAQKEENPDLIIIDIVKTYLSNEYSKDIDELNAYIEESLNKVTENSNDV	240	
Qy	241	RNLKFADEDLVRLYNQELVERVNLAAASDILIRISMLKBDGGVYLDVDDILPGQDPLFKS	300	
Db	241	RNFEEFKTGEVFNLYEYESVERVNLACASDILRAVLKNIIGVYLDVDMPLGHPDLFKD	300	
Qy	301	INKPDSI-TWTSNEMIKLEIMKYKEVYPGVTSKNFMDLEEVQSRSPESALSKSKDKEI	359	
Db	301	INKPDSVKTAVDWEEMOLEAMKHKYIPEYTSKHPTDLDEEVQSSPESVLASKSKSEI	360	
Qy	360	FLPLDDIKVSPLEVKTIAFANNSVINOALIISIKDQSCSDLVINOIKNRYKIITNLNAPSIN	419	
Db	361	FLPLGDIENVSPLEVKTIAPAKGSIINOALIISAKDQSCSDLLIKQIQNRYKIITNLDTLGPIS	420	
Qy	420	EGTDFTNTMKIFSDKLASISNEONMFMWIKITNYLKVGFAPDVRSITNLSPGVYTGAYQ	479	
Db	421	QGNDFNTMNFCSGLAIANEENISPIAKIGSVLRVGFYPEANTTITLSGPIIYAGAYK	480	
Qy	480	DLMLFKDNSNTNIIHLEPELANPFPPTKISQLTEQBITSLWSFNQARAKSQPBEYKKGYP	539	
Db	481	DLTLTFKMSIDTSLSELNRFPPKVNISQATEQEKNSLWQFNNEERAKIQFBEYKKNYF	540	
Qy	540	EGALGEDDNLDPAGNTVDKYVYSKTLSSMKTENKEYIHYIVOLQGDKISYEASCNLPS	599	
Db	541	EGALGEDDNLDFSQNTVTDKEYLLEKISSTKSEGGYVHYIVOLQGDKISYEAACNLFA	600	

A;Accession: S44273
A;Molecule type: DNA
A;Residues: 1-243;1204-2178 <HOA>
A;Cross-references: EMBL:223281
C;Genetics:
A;Gene: tcn-alpha
C;Superfamily: cpl repeat homology
C;Keywords: virulence factor
F;1980-1899/Domain: cpl repeat homology <CP12>

Query Match 27.4%; Score 1424.5; DB 2; Length 2178;
Best Local Similarity 32.7%; Pred No. 4.9e-46;
Matches 342; Conservative 211; Mismatches 437; Indels 55; Gaps 19

Qy 3 LVNKAQIQKMYVVFRIQEDEVAILNALBEEYH-NMSESSVYVKYLKLDINLLTDNYLN 61
Db :
2 LITREQLMKIASIPLKRPKEPYNILDALENFRNDIEGTSVKETYSKLSKLNELVDNYQT 61
Qy 62 TYKSGRGNALKPKPEYLTMTVEVLKNNSLP--VERKULHFIIWGGQINDTAINYINQWK 119
Db :
62 KYPSSGRNLALENFRDSLSELRILKNSRTSTIASKNLSPIWIGGPISDOSLEYNNMWK 121
Qy 120 DVNSDYTVKFVYDGSNAFLINTLKKTIIVESATNTLTSPRENLNDEPDYDNKYFKRMELI 179
Db :
122 MFNKDYNIRLPYDKNSLVNTLTAIQEESKVITIQNSQNILDTGYGHNFPSDRMKLI 181
Qy 180 YDKQHFDIVYKSI EE NPEPIIDNI IKTYLSNEYSKOLEALNKYIEBSLNKITANNGD 239
Db :
182 YRYKRELKMLYENKQNNS---VDDIINFISNFPKYDIGKLNQKENNNKMIAIGATD 238
Qy 240 IRNLEKFADELVRLYNQELVERWNLAASDIILRISMLEKGGVYDLVDILPGIQDFK 299
Db :
239 I-NTENILTNKLSYVYGELTQTNLAASDILRIALKKYGVYCDLDFLPGVNLSLFN 297
Qy 300 SINPKDSITNTSWEMIKLEATMKKEYIPGYTSKNFMDLSEVQSPESALSCKSEI 359
Db :
298 DISKPNMGDSNYWEAAIFEAANBEKGLMNPTPYKMEQVPSEIKERILLSFVN-HDIND 356
Qy 360 FLPLDDIKVSPLEV-----KI AFANNSVINQALISKDSYCSDLVINQIKRYKILNDL 414
Db :
357 ILPGDIKISQLEILLSRLKAATGCKTFSNAFIISNDSLTLNLLISQLENRYEILNSII 416
Qy 415 NP--SINEGT--FNNTTKIFSDKLASISNDNMFMFIKITNYLKGVPADVRSTINLSG 470
Db :
417 QEKPKICETYDSYINSVELVETTPKNLSMDGSSFYQQIIGLYSSGPKPEVNSTVPFSG 476
Qy 471 PGVVTGAYODLLMPKD-----STNIHLLEPELRNPFPPKTKISQLTEBITSLWSFNQ 524
Db :
477 PNIYSATCDTHYPIKNTFDMLSQNOBIPEAS-NNLYFSK-----THDEFKSLLARS 529
Qy 525 ARAKSQPEEYKKGVEFGEGALGDDNLDPAQNTVLDKDVYSKKI--LSSMKTRNKEYIHVIY 582
Db :
530 NIASEKEQKLITYIGRTLNYEDGLNFNKRKVTSELKVIEWNSTKIYENDVLMIL 589
Qy 583 QLQGDKISYEASCNLFSDXPYSILLYQXNIEG-SETAYYYVADAIEKIDKIRIPYQIS 641
Db :
590 QIQGDDISYESAVNVFGKNPKSIL---IQGVDDPANVPFYENGIVQSDNINNLSRFN 645
Qy 642 NKRNLIKLTFIGHGKSEFTDTFANLDVDSLSEI-----ETILNLAKADI SPKYIEINLL 696
Db :
646 DIKKIKLTLIGHGENVFNPKLFGGKTVMNDLTNIIKPLOHLLERREGVILKNKYLKINIL 705
Qy 697 GCNMFYSIYAEBTVPGKLLKIKDRVSELAMP-SISQDSITVSANOYEVRI NEEGREIL 755
Db :
706 GCYNFTPVKVDINSTFVGKLFNKIS---RDLPQKGFKNQGLEISANKYAIRINREGREV 762
Qy 756 DHSQGWINKBESI IKDISKEYI SFNPKNKII VKSYKYLHELSTLLOBRNNANSIDL 815
Db :
763 DYFCGWYSTDLIAEQISNKYVYVNEVENTLSARVEQLNKVAEPAKOI-----NSIIQT 817
Qy 816 EKKVMLETECEINVASNIDROI VEGRIEBAKWL TSDSI----NYIKNEPKLIESIDS LYD 871
Db :
818 TNNOELAQSLVNTYADLTITLYSELLKEDI PPEDLDNIOIKERIILNETLSRLHDPSNILD 877

Qy	872	LKHONGLDGSHFTSFPEDISKTEGNFRIRPINKETGNSIFITEKEIIFSEYATHLSKEISN	933
Dd	:	:::::	:::::
Db	878	FYQKNNISNMILFDSIIIEKQYNNVKLANKITGETSVIKTSDSLWNFTNKYKKIVDD	937
Qy	932	IKDTIPDNVAGLVKKYNLDAAHVBVTNLNSAFFQTQSLLIYNTTKESLSNLSVAMKVQVYA	991
Dd	:	:::::	:::::
Db	938	IKGIIVKDINGEFKKADFEIQNPQSLNSMLMOLLIDYKPYTEILLTNMTSLSKVQAYA	997
Qy	992	QLPFSTGLTNTDASKVVVELVSTALD	1016
Dd	:	:::::	:::::
Db	998	QIFOLSIGAIOEATEIVTIISDALN	1022
 RESULT 6 T00296			
C:Species: Escherichia coli plasmid pO157			
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 17-Nov-2000			
C:Accession: T00296; T42195			
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C. S.; Shinagawa, H.			
DNA Res. 5, 1-9, 1998			
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enteric			
A:Reference number: Z14127; MUID:98290540; PMID:9628576			
A:Accession: T00296			
A>Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-3169 <MAX>			
A:CROSS-references: EMBL:AB011549; NID:G4589740; PIDN:BAA31815.1; PID:G3337056			
A:Experimental source: strain EHEC O157:H7, substrain RIMD 0509952			
R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, P.R.			
Nucleic Acids Res. 26, 4196-4204, 1998			
A:Title: The complete DNA sequence and analysis of the large virulence plasmid			
A:Reference number: Z22068; MUID:98391744; PMID:9722640			
A:Accession: T42195			
A>Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-236, 'D', 238-1887, 'I', 1889-3169 <BUR>			
A:CROSS-references: EMBL:AF074613; PIDN:AAC70163.1			
A:Experimental source: strain EDL933; serotype O157:H7			
C:Genetics:			
A:Gene: toxB			
A:Genome: plasmid pO157			
A>Note: L7095			
C:Superfamily: Escherichia coli probable cytotoxin			
C:Keywords: cytotoxin			
 Query Match 7.9%; Score 408.5; DB 2; Length 3169; Best Local Similarity 20.7%; Pred. No. 1.2e-07; Matches 255; Conservative 200; Mismatches 449; Indels 325; Gaps 55			
Qy	14	YVKPRI---QEDEYVAILNAL--BEYHNMSSESSVVEKYKLKDKINLNFDNYLNTYKKSGR	68
Dd	:	:::::	:::::
Db	160	YIKIRKTGAEDQTTTQSLIINELLGVDRNTI-PQKISELNDIHSVENMQIKNSR	218
Qy	69	-----NKAKKPKX-----YLTMVELKKNLSLTPEVKNLFIWIGG	105
Dd	:	:::::	:::::
Db	219	KGIBELVKQGELLSSLINVNKGKOLSDNASKIINLGIBYQSHKV-DIEPFIHAVVAG	277
Qy	106	QINDTAINYINQWDVNSDYTVKFVYSNAF-----LINLKKT-----	144
Dd	:	:::::	:::::
Db	278	APPDNTSYIAFLNTYKDYTYLLWIDPNAFGAAKFGCILKNIAMYAIMELRTNPHLA	337
Qy	145	-----IVESATNTLE--SPRENLDNPEFY-----NKPYRKMEIIVDKQKH	186
Dd	:	:::::	:::::
Db	338	BENNEVLKIQNIQNETHETRETLERLKENRYKSLTSETKFKNVFFLESIMGWQDNY	397
Qy	187	IDYYSQIEENPEFI-----IDNIUK--TYLSNEYSKOLEALNKYIEESLNKITANNCG-	238
Dd	:	:::::	:::::
Db	398	FTYCISNGISNTDDISRDLFTFNWKUSPEVQNDPKSTVEKNKRDIDLKKNITISOKFGDR	457
Qy	239	----DIRNLEKFADEDLVRLYNQBELVRWNLAASDILRI-SMLKEDGGVLDVDILPGIQ	294

Db 458 FQLRDINTLESFKPKQDFFYQOQMLRNWYAAASQVRINILKEYGGIYTDILPAYS 517
Qy 295 PDLFKSINKPDSITNTSWEKILBAIMKYK--EYIPG--YTSKNFMDLDEEVORFESAL 350
Db 518 DKVSQIINE-KSDKRFFFDLKLRLIISLSIKLGEKYSIKH-DGLDETTLNQLNLT 575
Qy 351 SSKSDKSEIPLDLDIKVSPLEKIAFANNSVI-----NQALISIKDSY 394
Db 576 -SEIEK--LIIDDY-FKPVETKVVRDFTFKIFKRYQKWNTWIRGNWNNFMTLTKGSK 629
Qy 395 CSDLVINQIKNRYKILNDNLNPSINEGTDFTMTKIFSDKLASINE-----DN 443
Db 630 CIDFILSGQKQVLEL-QRIRDNISYNLFYTT-----EDLAKLNVAIGGIPAKKYLEH 683
Qy 444 MFMKIKITNLYKVGFPADVRSTINLSGP-----471
Db 684 GLF-----SEYRGDTTIPYVSTLNTSGPMIMROMKMYKSLGRIGEVHIKONKLSNVN 739
Qy 472 -GVYTCAYODLWMF---KONSTNIHLLEPE-----LRNFEPKTKISOLT---EQEIT 517
Db 740 LGVYASSNKONKSPWLNVPVSGINDITPDDESSWAVRNNDINKILFEKINCHVPEKLP 799
Qy 518 SLWSFNQARAKSQFEYKGYFEGALGEDDNLDFAGNTVLDKDYYSKILSSMKTRNKEY 577
Db 800 SLV-----YEIDRSRFFQGM---DNKSIRKHTVEINKDLI--KOINLLTSSNID 843
Qy 578 IHVIVLOQDKISYEASCNLFPSKDPYSSILYQKNIEGSTAYYYVADAEIKDYRIP 637
Db 844 VKLLIKL--DRELYAIASSKI-----DNPLARSIRTQLQANVYTSNTPENTINFI 895
Qy 638 YQISNKRN-----IKLFTIGHGKSEFNTDTFANLDVDS-----LSSEIE 676
Db 896 YDFYRKQDDLSAIAKL-----FSRNDADTKIIVWNSVMEKVFREVLS 941
Qy 677 TIINLAKADISPKYIEINLLGCNMFYSIYABETYPKLLKIKORVSLMPSISQDSI- 735
Db 942 CVLRSKKVD--SYINEN-----KKNLSKEDA--GALRDYAKLKMKELFMSLDDGKY 989
Qy 736 -TVSANQYVRNEEKREILDHSGKWINKESIIKDISKEYISNP-----782
Db 990 KIITTNAY---IKERDKL-----SGIYIENSIISGHESFDIIRSNQHEWGLSTVEOF 1041
Qy 783 KENKIIVSKYLHELSTLLQETIRNNANSSDILEKKVMTLCEINVAS-----NIDRQIV 837
Db 1042 KKEFEVVKSE-LSSAKSIIDDKNK-YITDPETKRNVLVHOLDSDIKERIAFLDISHYAY 1099
Qy 838 EGRIEAKNLTSDSIINYIKNEPKLIESI-----SDSLYDLKHQNG 877
Db 1100 PGSLEKQLS---GVFSDINIIEAYLLASVGSVSHGVVYVAPSDKLELLRRHT 1155
Qy 878 LDDSHFISFEDISKTENGFRIRINKETGNSI---FIETEKEIFSEYATHISKEISNIXD 934
Db 1156 KGNSEWI--EKITP-----YVYDILSDNVSNVLRPPLSEBQKILNDIKLEISKSVS--E 1206
Qy 935 TIFDNNVGLKVKVNLDAAEVNTLNSAFFIQLSIEYNTTK-----ESLSNLSV--- 983
Db 1207 QYPMKLTQKSSVIGIKYVDFDRYNNELFLSLPINQNLTLFPMYRYFEMLDIHIGIIE 1266
Qy 984 --AMKVQVYAQLFSTGLNTITDASKVVEL 1010
Db 1267 NKANREFIYKFSLSLNDFLINDERVNLN 1295

RESULT 7
DB1702
adherence factor TC0439 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: DB1702
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000

A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: DB1702
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-3225 <TET>
A:Cross-references: GB:AE002312; GB:AE002160; NID:g7190482; PIDN:AAF39293.1; PID:g719048
A:Experimental source: strain Nigg (MoPn)
C:GeneticS:
A:Gene: TC0439

Query Match 7.0%; Score 362; DB 2; Length 3225;
Best Local Similarity 19.8%; Pred. No. 6.7e-06;
Matches 254; Conservative 212; Mismatches 414; Indels 404; Gaps 63;

Qy 2 NLVNAQLOKMYVVKFRIQEDSEYVAIL-NALIEEYHNSSE-----SSVVEKYLKLDINN 54
Db 196 NILEKLSLTQ---EQQIKHNDLISNLKEAINRYSDLNKRKSRKQOSLLVROAKILD--E 249
Qy 55 LTDNYLNTYKSGRKNALKKKEYLTMVELKNNSLTPVEKNLHFIVIGQINDTAINY 114
Db 250 ILSQTKSTEERASNSVMTTIKEFTSHRV-----PVEKNHIGIWIAGSPPGTDEY 300
Qy 115 INQWKDVSNDYTVKPYVDSNAP-----LINTLTKTIVESATNNLTSPRENLDNPFVYNK 170
Db 301 IKLFLHTYPEFSFLPWVKTAAYGAAKFSSTLKRIFAFAVAVN---SLREATPEPVKQFVQ 356
Qy 171 FYRKMEIILYQKQHFIDYYSQIBENPEFIDNIKIYLSNEYSKDL-----218
Db 357 RYDK-LKKSYSYTSRDFDE--KQRLSEQVLVELYDNY-----NKFSKEIQSNFVDVLLHEM 407
Qy 219 -----EALNKY-----EALNKY-----IBESLANKI 232
Db 408 ITIQDSFFNYCOLKVGAITDETREYLEKVLKVEEDLSHYKETIKRKNSESIEKLKVEI 467
Qy 233 TANNNG-----NDIRNLEKFADEDLVRLYN--QELVERWNLAASDILIRISMLKEDGGYV 284
Db 468 NDSTGRERWVVKIRDLSKL--QDLTNSYNVETEMLLRWNYAAATDQLRWMLKEYGGIY 525
Qy 285 LDVDILPGIQDPLFKSIN-----KPDSTNT--SWEMIKL-----EAIMKYKEIPLYGT 331
Db 526 TDLTIMPOYSQDVLQKINDVGSGRFEHDKLRRTLTSLFAALKKLGSGKQTTVSFEAKKAMT 585
Qy 332 SKNFMDLDE-EVORSFESALSSKDSKISIFLPLDDIKV---SPLVKI---AFANNSVI 383
Db 586 LPTFTLQKQSISEIFKYLETETOAKKSIFQPMQDVTVVYRDPFPIQRYKHQWOTGNVRGL 645
Qy 384 NOALISLKDSYCSDLVINQIKNRYKILNDNLNPSINEGTDFTMTKIFSDKLASISNEDN 443
Db 646 NGLMMAHKDSAVVDARIAQRAAYDEM--ALRQNVVSGEFFRSL-----GDLEHVNREKN 699
Qy 444 M-MFMKIKITNLY-----KVGEPADVRSTINLSGQVYTGAYQDILLM--FKDNSTNI 491
Db 700 IGGYLA--NYLGGSLFFDFRQDSVIPGAISTLIGSGP-----DIIMDTMSYFTNL 749
Qy 492 -----HLEPELRN-----PEFPKTKISOLT-----EOBIT-----SLWSFNQA 525
Db 750 GPVGEDFLYEGKLGKAAFLGAYQAKTPKGELTYDLHLPLSIGANDVTPADASTWCETRQ 809
Qy 526 RAKSQF-----EYKKGYFEGALGEDD-----NLDFQA-----NTVLDK 559
Db 810 HCAAEALLLSDSISDSDEHPKIRRRERNPNDFSKLMSKEAQGLSDDFADLLPRFNLLIES 869
Qy 560 DYVSKKILSSMKTRNKEYIHYIVLOQDKISYEASCNLFPSKDPYSSILYQKNIESETAY 619
Db 870 SALDIHTLSAL---DRDIQHLFTKQV-----KDPVASVA-----900
Qy 620 YYYVADAIEIKEDKYR-IpyQISN-----KRNKILFTIGHGKSEF---658
Db 901 ----VFSLOQLAEMTRAIPIPIRNOVHILPEAQAHFEADWKKAIOILYLSHSHQTEVVIWY 957
Qy 659 ---NTDTFANLDVDSLSSEIETILNLAADISPKYIEINLLGCNMFYSIYABETYPGKL 715

Db 958 SSTHTQIVFGKOLLAAVERAAAKSL-MSDHPDPSLIT-----SYLKRYTQSHLGLV 1007
Qy 716 LLKIKDRVSELMPSISODSITVSANQYEVRIINEE---GKREILDHS-GKW---INKEE--- 766
Db 1008 TEFDOEDFELMVDIAEP---ELHKQLDKIEQVNSGLYSHVEHSLGWLKLSKEERKS 1064
Qy 767 ---SIIOISSKEYISFNPENKNIIVKSKY-----LHELSTLLQEIIRNNAN 809
Db 1065 KFLKILKETFEEREDSQQOQHTWFELRYKHQERKVDPAKKIQELITVQB-SQRVQ 1123
Qy 810 SSDID-----LEKKVMLTECEINVASNIDROIVEGRIBEAKNLITSINYIKNEFKLI 862
Db 1124 AQDIDITFAHKPFYQDLMKDQVAFEDISVITKYL-----LASDGVSGIITDPF 1173
Qy 863 ESISDSLYD-LKHQGLDSDHSPEDISKTEGPRIRF--INKETGNSIFITEK--- 915
Db 1174 PPPSKQLIDAMKQSLG-----EDFGELHYTLQWYDMLSKET-NSVTSEQAQKQLP 1223
Qy 916 ----BIFSEYATH-----ISKEISNIKOTIFDNV----- 940
Db 1224 QKLHEKLEGYTHDLLIPIDGVSALGLRPFSTEBGKVSDRVLTSLAPGVPNSASYAMTS 1283
Qy 941 -----NGKLKVKKNLDAHAVENTLNSAFTOSLLEYNTTKESLSNLSVAM-- 985
Db 1284 YLYGLFLITKDIQSGRLTHEI---LKERLQTYGGAYFI-----NESKIDVILLALS 1331
Qy 986 -KVQVY---AQLFSTGLNTITDAS 1005
Db 1332 KKAQISLIDAHKALTGFSFSSEAS 1355

RESULT 8
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28676; A45521
R:Singha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A>Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507; MUID:197077455; PMID:8920022
A:Accession: T28676
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <STN>
A:Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A>Title: Identification of the gene for a Plasmodium yoelii rhoetry protein. Multiple c
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KE>
A:Cross-references: GB:M34281

Query Match 6.7%; Score 350; DB 2; Length 2401;
Best Local Similarity 20.0%; Pred. No. 1.3e-05;
Matches 261; Conservative 202; Mismatches 391; Indels 454; Gaps 63;

Qy 27 ILNLEAEVHNMSESSVVEKY-----LKLKDIINNLTNLYTYKSGRNKALKKPK 76
Db 62 LINKVDSGNIIEENDINFDLSKPNPKLLEGLNGVFNKNWYK---NNKVLDPK 118
Qy 77 EYLTVMEVLKGNLSLTPVEKNLHFLWIGGQINDTAINYINQWKNVSDYT-VKFFVDSNA 135
Db 119 DTMKIILLI-----IQYNPEKGLDAMTKLNGEISQK 153
Qy 136 FLINT-LKTYIVESATNTLESFRENLDNPEFDYKFKRMEIYD----- 181
Db 154 FVINQIKQPKDKYDEKKEGFESL---ELAKN-WEKKKLEIITELKKNQNEETVOLDI 209
Qy 182 KQKHFDYVYKSIENPFIIDNIITKYLNS-----EY-----SKLEALNKYI 225

Db 210 KIRELIKQIKDIEBQK---IVNDLKLKLNKKIKETEKIEYIKAVDLKKIEKDNVYI 266
Qy 226 EESLN---KIT--ANNNDIRN-----LEKFAEDLVRVLYNQ--ELVERWNLAASDIL 272
Db 267 DELAKBPYQITKYIEKKEINEYNTIKSDFDKIYVGDIEQLYNEMFSVVOESNIEHNKT 326
Qy 273 RISMLK-BDGGVYLDV-----DILPGIOPDLFKSINK-- 303
Db 327 ELTLTKIDNVYNNIQNMETETVKSHLKNITNNKLSETILDIKIYGEITNELNKT 386
Qy 304 -----PDSINTSWEMIKLEA---IMKYKEYIPGY-----TSKNFMDLD 339
Db 387 EDFKNKEGLSKNIDEYAKENVQLNVYKSNILEIKKHYNDQINIDNIKEKAQYDQFK 446
Qy 340 EEVQSFESALSSKSKSEIFLPDDI----- 366
Db 447 EHMKTIPPENMKYQKPSIEIKIMKDEFUSKKNVNDPKVYKVESEHNKTELTNLIK 506
Qy 367 -KVSPLYKIAFAN-----NSVINQALISLSDSY-----C--SDLVINQI 403
Db 507 TEVSDEEK-KYENKFNDSKSLINETKKSIEEYQNTLKKVDDYIKVCLNTNELINC 565
Qy 404 KNRKILNDNLNPSINEGTDFNTWKIFSDKLASI----- 438
Db 566 HNKQTTKDKLQNKIKTIKETNSIDKIYTKFENILTDKTELETFTGLSLNNHESNNK 625
Qy 439 -----SNEDNMFM-----IKITNLYLKVGPAPDVRSTINLSGPGV 473
Db 626 ELLTYFYDLKANLGNKNENMLYKQFNEKEKAVEDIKQNV-----DINKIVSNIETI 678
Qy 474 YTGAVODLLMPKDNSTNIHLEPELRNPFEPKTKISQLSQTEQIEITSLWSFNQARAKSOPEE 533
Db 679 YTSIYN---INEDTENEIG-----KSELNLTNVLEKVKANVTNL-----NEIKE 720
Qy 534 YKGYFEGALGEDNDLDPAQNTVL--DKDYVSKILSSMKT-----RNKEYIHYIQLQ 586
Db 721 KLDYDFQDFGKEKNIKYDPENKIKNDITLQKIDKSITLTKKNSKSE--NHIDEIKG 778
Qy 587 --DKISYASCNLPSKDPYSILQKNEG-----SETAYYYVADAEIKEDIKYRIP 637
Db 779 QIDKLKKVPNKTMFNEPKKE---TEKKIENIVEKIDKKKNY-----KEIDK--LL 824
Qy 638 YQISNKRNIKLTFF-----IGHGKSEFNTDTFANL----- 666
Db 825 NEISKIENDKTSLEKKNINISYGS-----LGNLFQOIDEKKKAETHIKAMEAYID 878
Qy 667 DVDSL---SSRIETILNL---AKADISPKYIEINLILGCMFYSYSTY----- 706
Db 879 DLDNKKKSQIEKEMNIMNDIKMDI---HKEMKALNISHDDYKIYHTTSKNHEEKISDI 935
Qy 707 -----ABETYPGKLLLKIDRVSELMPSISODSITVSANQYEVRIINEEKRILL 755
Db 936 RKNSLKIIQDFSEESYINDIKKEKNVLE-----SQNNNT--DINQYLSKI--ENIYNIL 987
Qy 756 DHSGKWINKESSIIKDISSKEYISFNPENKNIIVKSKYLHSLTLLQEIIRNNANSSIDL 815
Db 988 K-----LNKIKKIDKV--KEYTDBIEKNNK-----KINAEUS-----NSEKIIQL 1027
Qy 816 EKKVMLTECEINVASNIDROIVEGRIBEAKNLTSIN-----YIK----- 856
Db 1028 KENSLKESQSKISTIDNVSECIKNITNLKTYIVNEKKNINITYFKNAEYQNVSIN 1087
Qy 857 -NEFKLIESISDSLYDLKHQGLD--DSHFISFEDISKTENGPRIRFKETG-NSIFIE 912
Db 1088 FNNIEMADTKSQYILNINKNNGTNTDYNIKELKEHKKSNVYK-----DEAGKNTQETK 1142
Qy 913 TEKEIFSEYATH-----SKEISNIKDTIFDNVNGKLKVKKNLDAHAVENT----- 958
Db 1143 KKNLEFKYEOEVTVLLNKKYVAVELUNKFKDT-KNYSQIIEKIK--DAHNTFTSQADKS 1199
Qy 959 ---LNSAFFIQLSLEYNTTKESLSNLSVAMKVQVYQALFSTGLNTITD 1003

Db 1200 EKKWNEIKNEQRIEDEVAKNNKSKAI-LDIQLSVEPFKIKFLKID 1246

RESULT 9
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28317
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: EMBL:AF063866; NID:94049647; PIDN:AAC97677.1; PID:94049717
C:Genetics:
A:Note: MSV156

Query Match 6.7%; Score 345.5; DB 2; Length 1127;
Best Local Similarity 21.9%; Pred. No. 7.3e-06;
Matches 248; Conservative 167; Mismatches 368; Indels 349; Gaps 54;

QY 9 LQKWVYKPRIODEYVAIILNALLEEYHNMSSESVVEKYLKLDINNLTDNLYNTYKKS-- 66
DB 61 VEKIFYHMQKFNQIYILQYIEYNEINKIKENKPFCK--NPL---YNYITYKKLY 115

QY 67 --GRNKALKKFEYLTWEVLEKNNSLTPVEKNLHPFIWGGQINDTAINYINQWQDVNSD 124
DB 116 IYDLVYEEKKORVELVIN--IEQK-----AVDK-----INDIKNNVNIHSDNETI 159

QY 125 YTVKVDYDNAPILNTLKTIVESAINTLESFRENLPDPFYNKRYKRWIEIYDKOK 184
DB 160 ITGK--ETLIDILNKLK-LVSSDEKQIEQIYKNNKEIEFKN----- 201

QY 185 HFIDYKQIEENPEFIIONIKYLSNEYSKQLEALNKYIEBSLKNITANNNDIRNLE 244
DB 202 -----INNVQK-----EINKQDELNLLDSEKKEFKKQ-----EELN 235

QY 245 KFADEDLVRLNOELVERWNLAASDILRISMLKEDGVYLDVILPFIQIPDLFKSINKP 304
DB 236 KTIDKK-----QEELIKKLN-----DKEINFNID-----EKQKLLDOI 268

QY 305 DSITNTSWEMIKLEATMK-YKEVPGYTSKNFMDLDEEVQSPESALSKSKSEIFLPL 363
DB 269 NSKINTLNENIK--GVMNLYTE-----TKNKISNLQNEILNK-DSTIKSLDEKQKLLDEL 320

QY 364 DDIKVSPLEVKIAFANNSVIN-QALI--SLKOSYCSDLVINOIKNRYKI-----LNDN 413
DB 321 DK-NINNTISLYNKSNTKITNQQLLESSLTFDNNANINELKSLIKLFDNDIOQLKND 379

QY 414 LNPISNEGTD-FNTTWKIFSDKLAS-----ISN-----EDNMFMKIKITNLYLKVGA 459
DB 380 IFEQNNKITDFFNNSTRIPKEKLDTEYKKIDDIKNNLQKLEESYKKIDEOYEEKKIN 439

QY 460 PDVRSITNLSGGPV-----YTGAYQDMLLPKDNSTNHLLEPELR 499
DB 440 KEYNDIIEKNNLQKLEBENKKIDEOYEEKKINKEYNIDIEELKNN-NLQKLEENK 497

QY 500 NFEFPKTKISLQTEQISITSLWSNQ-----ARAKSQPEYKKGVEFGALGED 546
DB 498 NINDKUTKLNDESN-TEL--FNKLNISDFDKDSREIAKLNTETEYEQRLKOLLENINKTN 554

QY 547 DNLDPAQNTV--LDKDYVSK-----KILSSMTKRNKEYIHYIVLOQDKISYEASCN 596
DB 555 ELMKLSDNKLSSLEQLYDSKNILDGIDKLYNSLKEKN-----DKI----- 595

QY 597 LFSKDPYSSILYQKNEGSETAYYYVVADEAIKEIDKYPVQISNKNIKUTFFIGHGKS 656
DB 596 ----DBYFS-----NIEKFDIYVNIENK-----FIGNLDS 621

QY 657 ----BFNTDTF---ANLDVDSLSSEIETILNLAKADISPKYTEINLNGCMFSSYSIYABE 709
DB 622 IINKIINDQFKEYINSKIDSKSNELSTWFD-----DIFNAKQIASITNINIS----- 672

QY 710 TYPGKLLKIKIDRVSELMPSISQDSITVSANQYVRINEEGKREILDHSGKWINKESII 769
DB 673 -----NKIKD-----LNEFIIS-NEDSSKELLDEIRKY-KQOFDKI 706

QY 770 KDISKEYISFNPKNKII--VKSXYLHELSTLLQEIQRNANSSDIDLEKKVMLTECEI- 826
DB 707 KDMAMTEVKSFEFTLQKIDISIKSN-INELTNAYDIINTKAN-----DLDDKLNNGSEPK 761

QY 827 ---NVASNIDROIIVGRITEEAKNLT-----SDSINYIKNEF--KLIE-----SISDS 868
DB 762 NLYNNASDLLDTIQKNDKVKQLNEYLEKKNQKQIEINDIVNNFIKELIKFNNTETNKS 821

QY 869 LYDLKHONGLDSPHFISFEDISKTENGPRIRFINKETGN-----SIFITEKEIFS---- 919
DB 822 LNELLTNDNDINDKI FLYKELNKISTNNLLKIKYNEIDNVNEKLSIVIENTLQFINSLSI 881

QY 920 -----EYATHIS-----KEISNIKOTIPDNVNGKLVKV 948
DB 882 EFNQGSITSHINFLNLTLAGINDVLNKLKIMADTTRGDTNRDEIKNQISSENISKQ 941

QY 949 NLDAAEHVNTLNSAFFIQSLIEYNTT-----KESLSNLSVAMKVQVY 990
DB 942 KFNKNEKDKKLLISFNDKLNKINISAGYTEYNNIEHECKLYIYAVSQEY 993

RESULT 10

T18427

Hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T18427

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z18935

A:Accession: T18427

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3724 <LAW>

A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1

C:Genetics:

A:Introns: 307/1; 1545/2

A:Note: C0335c

Query Match 6.4%; Score 330; DB 2; Length 3724;
Best Local Similarity 21.8%; Pred. No. 0.00013;
Matches 228; Conservative 176; Mismatches 351; Indels 292; Gaps 58;

QY 20 QDEYVAIILNALLEEYHNMSSESVVEKYLKLDINNLTDNLYNTYKSGNKKALKPKYVL 79
DB 618 EDDEYIHLK-----ENLKEDA-----NEYNNDKEN-----KNNKTEILSKNYL 658

QY 80 -----TMEVLEL--KNNSLTPVEKTLHPFIWGGQINDTAINYINQWQDVNSYTVKPYVD 132
DB 659 ENEKETLEELKLRGKNITFKDEKY-----NSLGEVIINEI-----QINEENKINDIQD 707

QY 133 SNAPLINTLKTIVESA--TNNLTLESFRENLPDPFYNKRYKRWIEIYDKOKFIDYV 190
DB 708 G-----NISKQKLIQSSRTNDTFNIKDISLD---DLEKEKVK-----KSKQFIDNL 753

QY 191 ----KSQIEENPEFIIDNIITY-----LSNEYSKOL-----BALNKYIEESLKNITANNNG 237
DB 754 VKADKNEISENINKICDNNINNIYDESINNIYDESINNIYDESINNIYDESINNIYDENI 813

QY 238 NDI--RNLEKFADEDLVRLNQ-----ELVERWNLAASDILRISMLKEDGVYLDV 287
DB 814 NNIDYENINNIYDENINNIYDEGINIKICDNTLENKNIKTNDIYVQV-----BENNESIE- 868

QY 288 DILPGIQDPLFKSINKPDSITNTSWEMIKLEAIKMYKEYIPGYTSKNFMDLDEEVQSP 347

Db	869	-----KNBLMISLAKD-----INNTYMPK-----ENVDIFINKIKR--E 901	Matches 255; Conservative 218; Mismatches 388; Indels 407; Gaps 65;
Qy	348	SALSSKSKSEIFLPDDIKVPLEVKIAFANNVSINOALISLKDSYSCSDLVINOIKRY 407	Qy 3 LVNKAQKQVYVYKPRIQEDVEVAITLNALEEVHNMSESVVEKYL--KLKDNNLTNDVYL 60
Db	902	SLL--KIDK-----NKK-----DNN---NDDEYMDNPYENDFIIN--HKM 935	Db 8 LVNLFVIVFV-YR-QYDKRSRALDKIKFVDLTKN--LEDFTIEDTKTEINDLAVD-M 63
Qy	408	KILNOLNP-SINEGDFNTWKIFSDKLAS---ISNEDNMF-MIKITNYLKVGFAPDV 462	Qy 61 NTKYKSGRNKALKKPEKYLTMVEVLKNNSLTPVEKNLHFHWIGQIINDTAINYNQWKD 120
Db	936	EITNKELDPLEINTQNEFIENLDIKKKYVNDHFNDADKMFVEMNKILN-----KDM 988	Db 64 BAYQRSS-IEIHKIEE--VQOKIKNSDPAEVEKKIAY-----HDSMLKOLDDEMTF 113
Qy	463	RTINLSGCVVTGAYQDILLMKDKNSTNHLLEPELRNFEPPKTKISQITQEBITSLWSF 522	Qy 121 VNSDYTVFVDSNAFLINTLTKTI-----VESATNTNLESF-----RENLDPEF- 166
Db	989	KKN-----KEQFFKTDFTFGLASHKIKKYNKGEKHDKNBEEKNILDYE 1035	Db 114 KVQENIQLOVDGK--IVDKLSKTLKGFNTQIDSVESNLNSVLEKDFKANKENLESIKTA 171
Qy	523	NOARAKSQPEYKKGKGFEGALCED-DNLOFAQNTVLDKDYV-----SKKIL--SSM 570	Qy 167 DYNKPYRKRMEIY--DKQKHFDYVKSQIBENPEFIIDNIK-----T 208
Db	1036	NOV-----YSLYSDHKIQDQDIHSIQTNICDENNIQINEENSKGVRISGTD 1087	Db 172 SWEKEDTNIKELVFKIDNLNKEISYKOLANIEBERKNDILVKNEKLDLDFSPDFLEKVE 231
Qy	571	KTRNKKEYIHYIYVLOQSD---KISYEASCNLFKDPYSSILYQKNIEGSETAYYYYVADAE 627	Qy 209 YLSNEYSKDLREAL-----NKY--IEESLNKITANNNDIRNLEKFADEDLVRLYNOELVE 261
Db	1088	ENKNDMENKNDMEKNDMEKNDIEKNDMEKND-----MEKKN-----DWE 1130	Db 232 FNIQKYSKEIESSFNFYENKYKLIENSIELIMESVKNKINEKEDF-----ILNRLNEELQN 287
Qy	628	IK-EDTKYRIPQISNKNRIKLTFFIGHGKSEFTDTPANLVDVDSLSSEIETILN-----680	Qy 262 RWNLAASDIL-----RISMLK---EDGCVLVDVI---LPGIQDLPKSIKPKPOSITWT 310
Db	1131	KKNDMEKKN---DMENENNME-----NKSDIENENRMEYKNNMKENKMIENNTIENN 1181	Db 288 KP-----KDLVYVDDRSKBIKDKLEDKLVLDNEISSMSSSFKNVYSRINSLE-----337
Qy	681	LAKADISPKYIEINLGCNMFYSYIABETYPGKLL---LKKDVSLSMPSISQDSIT 736	Qy 311 SWEMIKLEAIMKYKEYI PCYTSSK-----NFDMLDDEEVORSPESSALSS 352
Db	1182	MTENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENN 1236	Db 338 --ESWRIB-MGKYEBQVDDVDFKRSQVELNKNIEYEDKISQVDNIRREVELSLD 394
Qy	737	V---SANQYEVINEEGREILDHGSKWINKESIIKD-----ISSKEYISFNP 782	Qy 353 KSDKSE-----IFLPLD-----DIKV-----SPLVKI 375
Db	1237	IEDDAKQNVQNNQKVVHVKDKNLNNVD--IINDVLKSDDKFENNINSKE---INI 1292	Db 395 LNSKMESVQSGAIDFKIREDSDNGIYLEFPKGFADIEVPSFSGKGINQLKMWLESQ 454
Qy	783	KENKIIVKSK-----YLHELSTLLOEIRNANSDDILEKKWMLTECEINVASNIDRQIVE 838	Qy 376 APANNVSINOALISLKDSYSCDL---VINQIKRYKILNDNLN---PSNEGTFDFTT---427
Db	1293	KDFKVDKNDICLNLNELNK--QSYNSNTTEKCIDIH-----DFDYVENKO---VH 1339	Db 455 LDVDSN-IQEKLIKLNLDNLISNFEENGFRNNYSNLNDINAKYTALFESLSDSSSKPE 513
Qy	839	GRIEAKNLTSDSINYIKNEFKLIESDSLYDLKHQGLDSDSHFSPEDISKTEGPR 898	Qy 428 -----MKIFSDKLASISNEDNMMEIKITNYLVKVPADVRSTINLSPGVYTGAYQDL 481
Db	1340	DKIHEDRKEYCD-----ESKLYLPADDIKMR-----SPKINKGSKRENFLII 1384	Db 514 NOMESKYSFTDKLTAGMDEFSLMTGERF-----ETUSQEA-----TNNYQB- 555
Qy	899	RFINKETG---NSFIE-TEKIFSEYATHISKE-ISNIKDTIFDNVNGKLVKYNLDA 953	Qy 482 LMFKDNSTNHLLEPELRNFEPPKTKISQITQEBITSLWSFNQARAKSQPEYKKGKGFEG 541
Db	1385	SYMNEBIYFLIKKLVETIEPSIILKNMWHIEKQFLSPGKYINDNLK-----1432	Db 556 --FQD-----LNKLENEIESFY-----NMFKTQ-----578
Qy	954	HEVNTLNSAPFTQSLIENYNTTKESLN 980	Qy 542 ALGEDNDLDFQNTVLDKDYVSKKILSSMKTRNKYIHYIYVLOQGDKISY-----591
Db	1433	EGKCISKKIIDLMLVYNTTIISKN 1458	Db 579 ---ETLKVDFTNTSLINIKKEICKNIVEPRDRYDEVNIFVTQLEBSKLYQSKWQGMDSN 635

RESULT 11

G70163
 Hypothesis: protein B80512 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: G70163
 R:Praser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horat, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403685
 A:Accession: G70163
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2166 <KLE>
 A:Cross-references: GB:AE001153; GB:AE000783; NID:G2688419; PIDN:AA66876.1; PID:G268842
 A:Experimental source: strain B31

Query Match 6.1%; Score 318.5; DB 2; Length 2166;
 Best Local Similarity 20.1%; Pred. No. 0.00017;

QY 867 -----DSLYDLKH-----QNGLD-----DSHFISFEDISKTENGFRIRFIN 902
DB 964 FAAVLNNSBEFVKEVDSLLQDKETDIASQANIDITDLSLVKFNFINKEINGKYNEVIS 1023
QY 903 KETGNSIFITEKEIFSEYATHISKEISNIKOTIFDNVNGKLVKYNLDAAEHVNTLSA 962
DB 1024 NVRGYS-----ENISSKLENIEMHEIENLSRRLTRID-SLSKGM-----1063
QY 963 FFIQSLEIYNTKESISNLVAMKVQVY-AQLFSTGLNTITD--ASKVVELV-----S 1012
DB 1064 -----ENLQKLKESPDVSKYQVERFELVKOLTDGAEAKINKLVEIEQYK 1110
QY 1013 TALDETID 1020
DB 1111 SRLFEAID 1118
RESULT 12
GB1702
adherence factor TC0437 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: G81702
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
J.C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: G81702
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3255 <TEXT>
A:Cross-references: GB:AE002311; GB:AE002160; NID:g7190477; PID:AAF39291.1; PID:g719048
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0437
Query Match 6.1%; Score 316.5; DB 2; Length 3255;
Best Local Similarity 19.4%; Pred. No. 0.00035;
Matches 247; Conservative 197; Mismatches 472; Indels 359; Gaps 54;
QY 3 LVNAQLOKQVYVVFRIQDEYVAILNALAEYHNMSSESVVEKYLKDKDINNLTNLYNT 62
DB 200 IASRLKLTBEQIOKCHISLNN---LKAQORYLNIDESE--RNSRAGQDLLVKQSFPLN 254
QY 63 YKSGRNKALKPKFYLWELKNNSLTPVEKNLHFIWGGQINDTAINYNQWQVY 122
DB 255 ILENARWEESEEPKNVMSVIKTEF-NSHRVKIDSHPHGIWAGAPGTDVVIKTPQTY 313
QY 123 SDYTVKVVDSNAFLNTLTKTIVESATNTLTSPRENLDNDFDYNYKEY-----172
DB 314 BEPDFLFWDRQAFGAAKFSSILKTIAPDSSLKELRSITSQETQDPVKYDELOAKYQTS 373
QY 173 -----RKRMEIYKQKHID-----YKQSOIENPEF-----IID 203
DB 374 GRREERLEKOLKFLDXYKHLQSQIRTFDPLFKWVLSQDGFNYCLLKGITIDD 433
QY 204 NIKTYSNEYSKDLBALNKY-----IEESLNKITANNND-----IRNLEKFA 247
DB 434 QTRIAVLQDELHLPBEEIOYKOLIRANEQIKDVGKVTDLGKERVEIKNIRDLSMQ 493
QY 248 DEDVLRYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQPLDFKSINK-----303
DB 494 DRTNTYNEEMELLRWYPAASQIRMYMLKEGGIYITDLDMMQPYSPDVLQMINIGGD 553
QY 304 -----303
DB 554 RPEELPIRAVSDGVLRLANGESGITIDNIAQDIDISKLTRSDRTQLEKLLTDIENKQ 613
QY 304 -----PSGIT-----NTSEWMIKLEALMKYKVEIPGVTYSNQFDM 337
DB 614 SSGTSKFSLFQRMADPSVRDFMFIQRIHHKWSGTGMVNRGLNGLM-----MAHKGSA 665

QY 338 LDEEV--QRSFESALSSKSD---KSEIFLPLDDIKVSPLEVKIA-----F 377
DB 666 VDAVINGQORQAYRELKSLRETVLSGFEFKTLLDHLNHNKALIGGHLVEDFLGKSLFSEF 725
QY 378 ANNSVINQALISLSDSYCSDLVINQIKNRYKILNDNLNPSINEGTDFNTTKWIFSDKLAS 437
DB 736 RQDTIAGALSTLGIT-GPDLIVDEMKTYSRSLGP-IGKDVLEGKKLGN--KAPLGSYNK 781
QY 438 ISNEDNMMP--MIKITNVLKVGFPADVRSTINLSGPGVYTGAYQDLLMFKONSTHILLE 495
DB 782 IVKEGKQTFDMLHPVT---VG-ANDVTPADESTWCGIKQRCIAELL-FSDESK---FR 831
QY 496 PELRNFPEPK---TKISQLTEQEITSLWSFNQARAKS-----QFEYKKGFFEGALGED 546
DB 832 PET-----PKGISRTKIDKKT---FTKLWSEOSKALSPPELLDRFNE-----LIEA 874
QY 547 DNLDFAQNTVLDKO-YVSKKILSSMKTRNKYIHYIVQ---LOGDKISYEASCNLF---598
DB 875 KNFDLIKFTDVRDALYVLRNQVCCDASAAVFSLQQLAELLRCGTSFPIANQLNVFPDL 934
QY 599 SKDPYSSI-----LYQKNIEGSETAYY-----YVADAE---IKE--IDKVR 635
DB 935 NRNIASSLEKTIKLYQSHSOTETVIHSAISDRVMFLRDLMLATAERMLVKESIFSKEE 994
QY 636 IPYOISNKRNIKLTFIGHGKSEFNTDTPANLDVDSLSSEIEFTILNLAKADISPKVIEINL 695
DB 995 TP--UTTEIKLNSWAELOSKDNUGILSBEVDKLDVTSIMENSKLKKRVSALDSI 1052
QY 696 LGCNMFYSIYAEET-----YPGKLLLIKIDRVSELMPISISQDSITVSANOYEVRINBE 749
DB 1053 VAGYFY---LRLEETLSEWVRIPPDKLKRVRFMEGMDANQHPDERGGRWYSQLYDES 1109
QY 750 GK-----REILDHSGWINKBESIIKDSSKEYISFNPKENKIIVKSKYLHELSTLQ 802
DB 1110 FKRVTPAPVKKIQLAKKYLNEQRVHVLEID--DYLTKNPLFNR-----LHE-----1154
QY 803 EIRNANSSDIDLEKKVMLTCEINVASNIDRQIVEGRIEAKNLTSDSI-----NYI 855
DB 1155 ---EGYAFSDLTETRYMLAEYGIS-----GIFSEGNILPSPSARLVNIKITV 1200
QY 856 KNEFKLIESISDLSYDLKHQNG---LDDSHFISFED-----ISKTEN 894
DB 1201 GGDYHDMQDVLPIYDVLASGAADLTNERFAAIPESLRKNLEGLHGTDLTLPDVASVS 1260
QY 895 GPRIRFINKEGNSFIETEKEIFS---EYATHISKEISNIKDTIFD-----NVNGK 943
DB 1261 GWMGSF-GVENG---VESDHTMISIAFGFFNGASYSQHVLSALYIEHRIHILGSLTSD 1315
QY 944 LVKYNVLDAAHEVNTLNSAFFI---QSLLEYNTTKE--SLSNLSVAMKVQVY-AQLFS 995
DB 1316 LIKK-----ELESKGAGCFVHEERFDLSLKASSEKQVLSLUTEIHKLSUNOVHLAEVS 1368
QY 996 TGLNT-ITDASKVVE 1009
DB 1369 HLMNTALPGVGKIE 1383

RESULT 13

Ti8501
hypothetical protein C0760c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: Ti8501
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: Ti8501
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3394 <LAW>
A:Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331922; PIDN:CAB11140.1
C:Genetics:

A:Map position: 3
A:Note: C0760c

```

Query Match          6.1%; Score 315.5; DB 2; Length 3394;
Best Local Similarity 18.7%; Pred. No. 0.0004;
Matches 247; Conservative 212; Mismatches 422; Indels 439; Gaps 57;

QY 1  MNLVNAQLOQVYVYKRIQDEYVAIL-----NALEEYHNMSESVVEKYLKLDINNL 55
DB 1722  MWILKNKETENM--NIQKEEDYIKLIKDKTNIQNEVDL-----LEKYNEVVVKNNM 1773

QY 56  TDNYLNTYKSGRNKALKPKFYL-----TMEVLEKKNLSLTPVKRN-LHFTWIGQIND 109
DB 1774  LYNDNVN-----LLKEHKEEIFFLKENIKILOKNTYLMDFKQINNV-----DNN 1820

QY 110  TAINVINQWKNVSP-----YTVKVFY 131
DB 1821  LKLNLDQLFNQLDQKHLDTNQKHEQLKYDYIETIKERLKIETKKNQEKYIILOK 1880

QY 132  DSNAPLINTLKTTIVESATNTLTESFRENLDNPEFDYKFY-----YTVKVFY 172
DB 1881  DNN-LILNDFNSFTTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1939

QY 173  -----RKRMEIYDQ-KHFIDYKQ-----IEENPEFIID-NIITYL- 210
DB 1940  ELSNLEKIQLSDEKNRMKITTLEDKLPKNEKDKMKLQIIDDNNKNYMIQYKLNKLTNLD 1999

QY 211  -----SNEYSKOLEALNKYIEESLKNITANNNDIRNLEKFADE----- 249
DB 2000  MLSEENRMULLNKBEYEQIEQLNH--DHKLPISTNNNDIQIIEENEKQEQVDQYITTIN 2057

QY 250  -----DLVLYNQELVER-----WNLAAASDILRIS--MLKED--GGVLDVD- 288
DB 2058  BKDKIIVHLNLOIKLANQNEHRSRCDIFNVAHQDNINKDHVVVEDIMGDTNHDVNK 2117

QY 289  -----ILPGIQDLPKSNKPSINTNTSWEMIKLEAIM-----KYKE-----YIPG 329
DB 2118  NIDQGTNOHQINGTNOHQINGTNOHQINGTNOHQINGTNOHQINGTNOHQINGTNOHQING 2177

QY 330  YTSKFPDMLDE-----EVQSFESALSSKDSKSEIFLPLDDI-----KVSPL--- 371
DB 2178  YINENIDLTSELEKKNMULENYKNEKNEKE--EYKLNNDIDMLSNCKKLESIMME 2235

QY 372  EVKIAFANN-----SVINQALIS-LKDS-----YCSDL- 398
DB 2236  KYKIIIMNNIOEKDEIENLNKYNKLLDNLNNYSVVDKSVSCFEDSNIMSPSCNDIL 2295

QY 399  -VINOIKNRYK-----ILNDNLN--PSINEGTDFTNTTKIFS-DKLASISNEDMMFM 447
DB 2296  NVFNLSKSNKKVCTNMDCENMDSISSINNWNINNWNINNWNINNWNINNWNINNWNIN 2355

QY 448  IKITNLYKVGAPDVRSTINLSGPGYVTCAYQDILLMPKDNSTNIHLLEPELANPEPKTK 507
DB 2356  VDINYLNNL-----QLNKDNDNIIIIKFNILKFLGSCYLYIINRNLEIQMLKNQ 2409

QY 508  ISOLTEOEITLSWSP--NOARAKSOFEEYKGYFEGALGEDNLDFAQNTVLD-KDYVSK 564
DB 2410  ILSL-BESIKSLNEFINLNKNEKNEKELIKINNFEILKLNKLNQDNESCINLNLYLK 2468

QY 565  -KILSSMTRN-----KEYIHYIVQLQGDKISYEASCNLFSKDPYSILYQKNIEG---SE 616
DB 2469  NEELKNVKNPKYKGYIHLIQ-----QSNVFCIKPKFNENKIDQSIINKLLYIK 2522

QY 617  TAYYVVYVADAEIKEDIKYRIPQISNKNRIKLTFTGHGKSEFNTDTFANLDVDSLSSEIE 676
DB 2523  KSFDFMYDVSIOE-----IRENKNIII-----NODFLTDEY-----FKHIG 2559

QY 677  TILNLAKADISPKYIEI-----NLLQ-----CNMFSYS----- 704
DB 2560  TPTKTCNVLIQRYGISILKDTNDFPIQNKQSQGNQGNQGNQGNQGNQGNQGNQGNQGNQ 2619

QY 705  -----IYAETYPGKLLI----- 717

```

```

DB 2620  QIFDGTENVQOSLQNEEDYVNVNEEMYTDKMDLNNMNGDDDDDDDDDDDDNNNNNNNN 2679
QY 718  -----KIDRVSELMPSISQDSITVSANOYEVRIINECKREILDHSGKWINKESIIKDI 772
DB 2680  NNNNNMGDDHNLVAFNNHLLTNGVSKDQINNETLRY-----EENIQNI 2729
QY 773  SSKEYISFN--PKENKIIIVKSKY-----LHELSTLLOIRNANSSDIDLEKK 818
DB 2730  YTNDDVNDQVNIENKILIKQDQIINNDELKNEHNNLIRLINESIAH-----NLENV 2785
QY 819  VMLTECEINVASINDROIVEGRIBEAKNLTSDS-----INVIKNEPKLIESI 865
DB 2786  YVQNDANNLINDNKKSETLTIVDEKONVSNESSKCDKKKKNEDITQAKNENYFVSTH 2845
QY 866  SDSLYDLKHQGLDSDHSPISPEDISKTEGFRIRPINKETGNSIFIETEKEIFSEVATHI 925
DB 2846  YDNDDINKNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNND 2892
QY 926  SKEISNIKDTTIFDNNVNGKLV-----KKVNLDAAHEVNTLSAFPQISLIEN 972
DB 2893  NINDDNNNDNNNDNNNGFVCELSSNINDFNILNVKDNFGINKSN--FSTNLSEYN 2950

RESULT 14
JC6009
surface-located membrane protein lmp3 precursor - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: JG6009
J. Bacteriol. 178, 2775-2784, 1996
A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene system
A:Reference number: JG6009; MUID:96213016; PMID:8631664
A:Accession: JG6009
A:Molecule type: DNA
A:Residues: 1-1302 <LAD>
A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CAA64858.1; PID:g1197336
C:Genetics:
A:Gene: lmp3
A:Genetic code: SG3
C:Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homology
C:Keywords: duplication; membrane protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-1302/Product: surface-located membrane protein lmp3 #status predicted <MAT>
F:957-992/Domain: tetratricopeptide repeat homology <T1>
F:993-1026/Domain: tetratricopeptide repeat homology <T2>
F:1089-1120/Domain: tetratricopeptide repeat homology <T3>
F:1154-1190/Domain: tetratricopeptide repeat homology <T4>

Query Match          6.0%; Score 311; DB 1; Length 1302;
Best Local Similarity 20.1%; Pred. No. 0.00017;
Matches 234; Conservative 204; Mismatches 396; Indels 330; Gaps 53;

QY 3  LVNKAQLOQVYVYK-FRIQDEYVAILNALBEEYH-----NMSSSVVE--- 44
DB 169  LLNKINAEERLQSKIFNEKKQELKRVLD-LEDTKEVDFTYKQVFIETINETSIEDIK 227

QY 45  -KYLKL-KDINLTDNYLNTYKSGRNKALKPKFYLTMEVLEKKNLSLTPVKRNLFHW 102
DB 228  NKIIIEVKATSSLTSKIILNT-----KQELQEF-ENIKKLDQDFINTKLNDAKYQ----- 276

QY 103  IGGQINDTAIVINQWKNVSDYTVKFPVYVDSNAPLINTLKTTIVESAT--NNTLESPE 159
DB 277  ---SLQKALDKINSINGINKNSTIKEIKAGONALIKAKEEAGLEKEKLDGQNIKDTLKE 333

QY 160  NLNDPEFDYKRYKRMEEIYDKQKHFDYKYSQIEENPEFIIDNIKITLSNEYSKOLB 219
DB 334  TINNAK-----EPKLLIDNDQKIVD--LKSNDLNEISKARQ 368

QY 220  ALNKYIEESLKNITANNNDIRNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKE 279
DB 369  SLSK-----DKMESANDLLNTKLBYKEILKNFNOEKAFFN-----ELEQTRK 414

```


Thu Nov 6 08:28:34 2003

us-09-126-816b-6_copy_1_1020.rag

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:25:02 ; Search time 30.75 Seconds
(without alignments)
5265.077 Million cell updates/sec

Title: US-09-126-816B-6_COPY_1_1020

Perfect score: 5192

Sequence: 1 MNLVNAQLQKVVYKFRIQ.....ITDASKVVELSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	4086	78.7	2366	17 AAR95011	C. difficile toxin
2	4083	78.6	19 AAW68388		Clostridium diffi
3	2783	53.6	2710	17 AAR95016	C. difficile toxin
4	2783	53.6	2710	19 AAW68387	Clostridium diffi
5	2780	53.5	546	20 AAV33700	C. sordellii letha
6	293.5	5.7	1254	11 AAR07503	Merozoite apical-en
7	293.5	5.7	1254	18 AAW24575	Merozoite apical-e
8	286.5	5.5	1979	21 AAB18171	Plasmodium falcipa
9	280	5.4	1516	21 AAB18195	Plasmodium falcipa

10	272	5.2	980	21 AAB18294	Plasmodium falcipa
11	266.5	5.1	1121	21 AAB18241	Plasmodium falcipa
12	261.5	5.0	2184	22 AAE00425	P. falciparum telo
13	258.5	5.0	1712	21 AAB18205	Plasmodium falcipa
14	256	4.9	1090	23 ABB54799	Lactococcus lactis
15	256	4.9	1948	23 ABP73774	Candida albicans e
16	255	4.9	1817	21 AAB18301	Plasmodium falcipa
17	255	4.9	2485	21 AAB18172	Plasmodium falcipa
18	249	4.8	1639	19 AAW54145	P. falciparum synt
19	249	4.8	1639	23 AAE29345	Plasmodium falcipa
20	247.5	4.8	10182	23 ABP38314	Staphylococcus epi
21	246	4.7	1881	23 ABP73809	Candida albicans e
22	245.5	4.7	5024	22 AAG82935	S. epidermidis ope
23	242.5	4.7	1346	21 AAB18236	Plasmodium falcipa
24	241.5	4.7	1089	23 ABB09491	Ampev first DNA po
25	238	4.6	1477	20 AAY19981	B. burgdorferi ant
26	238	4.6	1494	20 AAY19980	B. burgdorferi ant
27	237.5	4.6	3433	18 AAW22017	Utrrophin. Homo sa
28	237	4.6	1654	6 AAP50777	Sequence of the P1
29	236.5	4.6	1308	21 AAB18167	Plasmodium falcipa
30	236.5	4.6	1802	21 AAB18217	Plasmodium falcipa
31	235	4.5	3210	24 ABU07438	Protein differenti
32	235	4.5	3248	17 AAR99795	Kinetochore protei
33	232	4.5	3973	21 AAB18253	Plasmodium falcipa
34	229.5	4.4	2681	24 ABJ19025	Pathogen specific
35	224.5	4.3	1151	23 ABP38397	Staphylococcus epi
36	224	4.3	1270	23 ABP73290	Candida albicans e
37	222.5	4.3	2663	22 AAM39097	Human polypeptide
38	221.5	4.3	1558	21 AAB18324	Plasmodium falcipa
39	220.5	4.2	2295	21 AAB18180	Plasmodium falcipa
40	220	4.2	1188	21 AAB18183	Plasmodium falcipa
41	220	4.2	3696	23 ABP40235	Staphylococcus epi
42	219	4.2	1714	21 AAB18275	Plasmodium falcipa
43	219	4.2	2496	21 AAB18222	Plasmodium falcipa
44	218.5	4.2	1145	22 AAG82169	S. epidermidis ope
45	218.5	4.2	1430	20 AAY31949	Plasmodium falcipa

ALIGNMENTS

RESULT 1
AAR95011
ID AAR95011 standard; Protein; 2366 AA.
XX
AC AAR95011;
XX
DT 08-JUL-1996 (first entry)
XX
DE C. difficile toxin B.
XX
KW Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin;
XX diarrhoea; therapy; diagnosis; vaccine.
XX
OS Clostridium difficile VPI strain 10463 (ATCC 10463).
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US13737.
XX
PR 07-JUN-1995; 95US-0480604.
PR 24-OCT-1994; 94US-0329154.
PR 16-MAR-1995; 95US-0405496.
PR 14-APR-1995; 95US-0422711.
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
PI Williams JA;
XX
DR WPI; 1996-230603/23.

DR N-PSDB; AAT29247.
XX Fusion proteins comprising non-toxin protein and part of toxin -
PT useful to form anti-toxins against Clostridium botulinum type A, and
PT C. difficile type toxins, and to treat C. difficile intoxication,
PT partic. diarrhoea
XX
PS Claim 36; Page 313-323; 434pp; English.
XX
CC Clostridium difficile VPI strain 10463 toxin B (AAR95011), the
CC product of the toxin B gene (AAT29247), is a cytotoxin associated
CC with diarrhoeic disease. It can be obtd. by expression in
CC transformed E. coli hosts of portions of DNA that together cover the
CC entire toxin B gene. Toxin B, and portions of it (see also AAR95012-13,
CC AAR7371-72 and AAR95018), pref. expressed as fusions to polyhistidine
CC affinity tags or maltose binding protein, are used to raise avian
CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.
XX
SQ Sequence 2366 AA;
Query Match 78.7%; Score 4086; DB 17; Length 2366;
Best Local Similarity 77.4%; Pred. No. 1.5e-213;
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;
QY 1 MSLVNRKALQKQVYKFRIOEDYVAIINALEEYHNSSSVVEKYLKLDINNLTDNYL 60
DB 1 MSLVNRKQLEKMANVFRIOEDYVAIINALEEYHNSSSVVEKYLKLDINSLDIYI 60
QY 61 NTYKSGRNKALKKFKYELTMEVLELKNLSLTPVERKNLHPITGGQINDTAINYNQWD 120
DB 61 DTYKSGRNKALKKFKYELTMEVLELKNLSLTPVERKNLHPITGGQINDTAINYNQWD 120
QY 121 VNSDYTVKPYVDSNAFLINTLTKTIVESATNTLESFRENLPNDPEYNNFKRMEIY 180
DB 121 VNSDYNNVYFDSNAFLINTLTKTIVESATNTLESFRENLPNDPEYNNFKRMEIY 180
QY 181 DKQKHIDYKSOIENPEFIIDNIKTLYSNEYSKOLALNKYIBESLNKITTANNNDI 240
DB 181 DKQKNINYKAQREENPELLIIDIVKTYLSNEYSKEIDELNTYIBESLNKITTQNSNDV 240
QY 241 RNLEKFADEDLVRLNQELVERNLAAASDILRISMLKEDGGVYLDVLPQDPLFKS 300
DB 241 RNLEKFADEDLVRLNQELVERNLAAASDILRISMLKEDGGVYLDVLPQDPLFKS 300
QY 301 INKPDSITNTSWEIKLEAMTKYKEVTPGYTSKNFDMLEEVQSFESALSSKSKSEIF 360
DB 301 IEKPSVYVDFWEMTKLEAMTKYKEVTPGYTSKNFDMLEEVQSFESALSSKSKSEIF 360
QY 361 LPDDDIKVSPLVKIAFANNSVINQALISLKQSDLVINQIKNRYKIINDNLPSINE 420
DB 361 SSGDMEASPLEVKIAFNSKGIINQGLISVKQSYCSNLIVKQLENYKILNLSNPAISE 420
QY 421 GTDFNTMTKIFSPLKASINEDNMFMKITYNLKVGFPADVRSTNLGSGVYVYQAYQD 480
DB 421 DNDFTNTTTFIDISMAEANADNGRFMWELGKLVGRVFPDVKTITNLGPEAYAAAYQD 480
QY 481 LMFKDNSTNIHLEELRNFEPKTKISQLEITSLWSFNQAKSQFEYKKGYE 540
DB 481 LMFKDGSNMNIHLEADLRFESKTNISQSTQEMASLWSFDDAKAKQFEYKKNYFE 540
QY 541 GALGEDNDLDFQNTVLDKDYVSKKTLSSMKTNRKEYIHYIVQLQGDKISYEASCNLFSG 600
DB 541 GSGEDNDLDFSNIVVDKEYLEKISLARSSEYGIHYIVQLQGDKISYEACNLFPAK 600
QY 601 DPVSSILYQKNIGSETAYVYVADAEIKIDKIRIPYQISNKRNTKLFIFGHGKSEFNT 660
DB 601 TPVDSVLFOKNIEDSEIAYVYNNPGDGEIQEIDKIPSIISDRPKIKLAFIGHGKDEFNT 660
QY 661 DTEANLDVLSSEIETILNLAKADISPKYIEINLGCNMFYSIYAEETYPCKLILKIK 720
DB 661 DIFAGFDVSLSTEIAAIDLAKEDISPKSIEINLGCNMFYSINVEETYPCKLILKIK 720
QY 721 DRVSELMPISQDSITVSANQYEVRIINEEGKREILDHSGKWKINKESIIKDISKEYISF 780

DB 721 DKISELMPISQDSITVSANQYEVRIINEEGREILDHSGWKINKESIIKDISKEYISF 780
QY 781 NPKENKIIVKSKYLHELSTLQEIERNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840
DB 781 NPKENKIIVKSKNLPESLTLQEIERNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840
QY 841 IESEAKNLTSDSINYIKNBFKLTESISDSLYDLKHONGLDSDHFI SPEDISKTENGPRIFR 900
DB 841 IESEAKNLTSDSINYIKNBFKLTESISDALCDLQOQNELEDHFI SPEDISKTENGPRIFR 900
QY 901 INKETGNSIFITEKEIFESEYATHISKEISNIKDTIFONVNGKLVKKNVLDAAHEVNTLN 960
DB 901 INKETGESIFVETETIFSEYANHITELSKIKGIFDTVNGKLVKKNVLDTHVNTLN 960
QY 961 SAFFQSLIEYNTTESISNLSNVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAFQISLIEYNTTESISNLSNVAMKVQVYAOQFSTGLNTITDAAKVVVELVSTALDETID 1020
RESULT 2
AAW68388
ID AAW68388 standard; Protein; 2366 AA.
XX
XX AAW68388;
XX AC
XX 07-DEC-1998 (first entry)
XX
XX Clostridium difficile toxin B.
XX DE
XX Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen;
XX KW pseudomembranous enterocolitis.
XX
XX Clostridium difficile.
XX OS
XX W09808540-A1.
XX PN
XX 05-MAR-1998.
XX PD
XX 28-AUG-1997; 97WO-US15394.
XX PF
XX 28-AUG-1996; 96US-0704159.
XX PR
XX (OPHI-) OPHIDIAN PHARM INC.
XX PA
XX Thalley BS, Williams JA;
XX PI
XX WPI; 1998-230234/20.
XX DR N-PSDB; AAV30561.
XX
XX Host cell containing recombinant expression vector encoding
XX Clostridium botulinum type B or E toxin - useful to treat humans
XX and other animals at risk of intoxication with clostridial toxin
XX
XX Example 18; Page 241-249; 428pp; English.
XX
XX This is the amino acid sequence of Clostridium difficile toxin B,
XX deduced from the coding region (see AAV30561) of the toxin B gene.
XX Fragments of the toxin B gene have been cloned into various
XX prokaryotic expression systems, and assessed for the ability to
XX express recombinant toxin B protein in E. coli. It would be
XX advantageous to use simple and inexpensive prokaryotic expression
XX systems to produce and purify high levels of recombinant toxin B
XX for immunisation purposes. The invention specifically relates to
XX recombinant proteins derived from Clostridium botulinum toxins
XX (see AAW68389-400) and their use as immunogens for the production of
XX vaccines and antitoxins.
XX
SQ Sequence 2366 AA;
Query Match 78.6%; Score 4083; DB 19; Length 2366;
Best Local Similarity 77.4%; Pred. No. 2.1e-213;
Matches 789; Conservative 106; Mismatches 125; Indels 0; Gaps 0;

Qy	1	MNLVNKAQLQGVVYKFRIOEDYVAINALBEEYHNMSSESVVEKYLKLDINNLTNYL	60
Db	1	MSLVNRKOLEKQWNRFTQDEYVALDAALEEYHNMSNTVVEKYLKLDINSLTDIYI	60
Qy	61	NTYKCSGRNKALKFKPEYLTMEVLELKNLSITPVEKNLHFIWIGQINDTAINYINQWKD	120
Db	61	DTYKCSGRNKALKFKPEYLTVEVLEKNNLITPVEKNLHFVWIGQINDTAINYINQWKD	120
Qy	121	VNSDYTVKVVVDSNAFLINTLTKTIVESATNNTLESRENLDNPEYDNKPYFKRMRBIY	180
Db	121	VNSDYNVNVFYDSNAFLINTLTKTIVESANDTLESRENLDNPRPDYNAFFKRWQIY	180
Qy	181	DKQKHFDYKQSIIEENPEFIIDNIKITYSNEYSKOLEALNKYIESSLNKITANNNGDI	240
Db	181	DKQKNFINYKQAREENPELLIIDIVKTYLSNEYSKIDEIANTYIESSLNKITQNSGNDV	240
Qy	241	RNLEKFADEDLVRILYNQELVERNWLAAASDILRISMLKEQGVVLDVLDIILPGTODLPKS	300
Db	241	RNFEFKNGSEFNLYEQELVERNWLAAASDILRISALKEIGMVYLDVMDLPGTODLPES	300
Qy	301	INKPDSITNTSWEMIKLEAMTKYKEYIPGYTSKNFMDLBEVQRFSEFALSXSIXDSEIF	360
Db	301	IEPSSVTVDVFWEMTKLEAMTKYKEYIPETSEHFDMLDEVOSSFESVLASKDSEIF	360
Qy	361	LPLODIKIVPLEVKIAPANNVINOALISLKDSYCSDLVINQIKNRYKILNDNLNPSINE	420
Db	361	SSLGDMEASPLEVKIAPANSKGIINOGLISVKDSYCSNLIKVQIENRYKILNLSNLPaise	420
Qy	421	GTDPNTWKIPSKLASHEDNMFMKITYNLYKGFADPRSTINLSPGVVTGYQD	480
Db	421	DNDFTNTTNTFIDSIMAEANADNGRFMELGKYLVRGVFPDVKTITINLSGPEAYAAAYQD	480
Qy	481	LLMPKONSTNHLLEPELRPEPEPKTISOLTEQEITSLASFNOARAKSOFEEYKGYE	540
Db	481	LLMPKESGMNHLLEADLRNPEISKTWNISQTEQEMASLWFDARAKAQFEYKRYFE	540
Qy	541	GALGEDNLDPAQNTVLDKDYVSKKILSSMKTRNKEYIHIVOLQGDKISYEASCNLFSK	600
Db	541	GSLGEDNLDPSQNIIVDKEYLLEKISSLARSSERYIHIVOLQGDKISYEAAACNLPak	600
Qy	601	DPYSIILYKQNIIEGSEYAYYYYADABIKEDKTRIPIQISNGKNIKLITFIGHKSEFNT	660
Db	601	TPYDSVLFOKNIEDESEIAYYYPNGDGBIQEIDKYKIPSIISDRPKIKLITFIGHKDEFT	660
Qy	661	DTFANLDVDSLSSEITILNLAADISPKYTEINLGCNMFYSIYABEYTPGKLLKIK	720
Db	661	DIFAGFDVDSLUSTEIEAIDLAKEDISPKSTEINLGCNMFYSINVEEYTPGKLLKVK	720
Qy	721	DRVSELMPSISQDSITVANSQYEVRIINEGKRETLDSGWINKKEESIIKDISKEYISF	780
Db	721	DKISELMPSISQDSIIVANSQYEVRIINSEGRRELLDHSWINKKEESIIKDISKEYISF	780
Qy	781	NPKENKIIVSKYLUHELSTLLOEIRNNANSDDIDLEKKVMLTECEINVASNIDRQIVEGR	840
Db	781	NPKENKITVSKNLPESLTLLOEIRNNNSSDIELEEKVMLTECEINVISNIDTQIVEER	840
Qy	841	IEEAKNLTSDSINYIKNEFKLIESISDSLYDLKHONGLDOSHFTSPEDISKTENGFPIRF	900
Db	841	IEEAKNLTSDSINYIKDEFKLIESISDALCDLKOQNELEDSHFTSPEDISDEGFSIRF	900
Qy	901	INKETGNSPIETEKEKIFSEYATHISKEISNIDTIIDFNVNGKLVKVKNIIDAAHEVNTLN	960
Db	901	INKETGESIPVETEKTIKIFSEYANHITSEISKIGTIIDFTVNGKLVKVKNIIDTHEVNTLN	960
Qy	961	SAPFIQSLIEYNTTKESLNSLVAMKVVQVYAAQLFSTGLNTITDASKVVVLVSTALDETID	1020
Db	961	AAPFIQSLIEYNSKESLNSLVAMKVVQVYAAQLFSTGLNTITDAAKVVVLVSTALDETID	1020

RESULT 3
 AAR95016
 ID AAR95016 standard; Protein: 2710 AA.

RESULT 3
AAR95016
ID AAR

Qy	541	GALGEDDNLDPQANTVLDDKY-VSKKTLSS--MKTNRKEVYHYIYVLOQGDKIYSEASCNL	597
Db	540	GSUSEDGDFNKNVNTALDKNYLLKNSPNNVEAGSKVYHYIYIQGDDIYSYEATCNL	599
Qy	598	FSKDPYSGIILYQKNIEGSETAYYYYVAD--AEIKEIDKYRIPYQISNKRNIKLTFIGHGK	655
Db	600	FSKNPKNSIIIQNM--NESAKSYFLSDGSEILELNKYRIPERLKNKEKVVTFIGHGK	657
Qy	656	SERNUTDFANLDVDSLSEIETILNLAKADISPKYIEINLLGCMNPFYSIYABETYPGKL	715
Db	658	DEFNTSEFALSDVDSLSEISSFDTYTKJDISPKNVEVNLGCMNPFYSDFNVETYPGKL	717
Qy	716	LLKIKDRVSELMPSISODSITVSANQYEVRAINERBEKREILDHSGKWINKEESIYKIOISSK	775
Db	718	LLSITDKITSTLPDVKNKSITIGANQYEVRAINSEGRKELLAHSGKWINKEEAIMDSLSSK	777
Qy	776	EYISFNPENKNIIVKSKYLHELSTLLQEIERNNANSIDIEKKVMLTECEINVASNTDRQ	835
Db	778	EYIFPDSIDNKLKAKSNIPGLASISSEDIKTLLLDASVSPDTKILNNLKNITSESSIGDY	837
Qy	836	IVEGRIEAEKNLTSDSINYIKNEPKLTESISDSLYDLKHONGLDSDHFI SFEDISKTENG	895
Db	838	IYYEKLFPVKNIIHNSIDDLIDENFNLENVSDLEYELKLNLDKYLII SFEDISKNST	897
Qy	896	FRIRFINKEGNSIFITEKEIFSEYATHIISKEISNIKOTIFDNVNGKLVKKNVLDAAHE	955
Db	898	YSVRFINKSGESYVTEKEIFSKYSEHITKEISTIKNSIITDVNGNLLDNIIQLDHTSQ	957
Qy	956	VNTLNSAFFTQSLEIYNTTKESINLSVAMKVQVYAOLFSTGLNTIITDASKVVELVETAL	1015
Db	958	VNTLNAAFFTQSLIDYSSNKKVDNLSTSVKQVLYAQLFSTGLNTIITDASKVVELVETAL	1017
Qy	1016	DEIID 1020	
Db	1018	NDIIN 1022	

RESULT 5	
AAV33700	
ID	AAV33700 standard; protein; 546 AA.
XX	
XX	
AC	AAV33700;
XX	
DT	18-JAN-2000 (first entry)
XX	
XX	
XX	C. sordellii lethal toxin protein fragment.
XX	
KW	Lethal toxin; immunotoxin; antitumor; glucosyl transferase; glycosylate;
KW	CPase; Ras; epidermal growth factor stimulated MAP-kinase; inhibition;
KW	signalling pathway; cell-specific toxin; treatment; cancer.
XX	
OS	Clostridium sordellii.
XX	
PN	DE19802569-A1.
XX	
PD	09-SEP-1999.
XX	
XX	23-JAN-1998; 98DE-1002569.
XX	
PR	23-JAN-1998; 98DE-1002569.
XX	
PA	(UYFR-) UNIV FREIBURG ALBERT-LUDWIGS.
XX	
PI	Aktories K, Hofmann F;
XX	
DR	WPI; 1999-509323/43.
DR	N-PSDB; AA23800.
XX	
PT	New fragment of the lethal toxin from Clostridium bacterium, useful for
PT	treating cancer -
XX	
PS	Claim 1; Page 7-9; 14pp; German.
XX	

This invention describes a novel fragment (I) of the lethal toxin (LT) of *Clostridium sordellii* which has antitumor activity. (I) is a glucosyl transferase that glycosylates, and thus inactivates, GTP/guanine triphosphatases, particularly Ras (an oncogenic product overexpressed in many tumors), resulting in inhibition of epidermal growth factor stimulated MAP-kinase signalling pathways. (I), particularly in the form of immunotoxins, are used as cell-specific toxins, particularly for treating cancer. When included in immunotoxins, (I) can be targeted to selected cells. Compared with the complete LT, (I) is smaller, so enters cells more easily, resulting in greater toxicity in the cytosol, is less likely to induce formation of (neutralizing) antibodies, and is more active than the holotoxin. This sequence represents the lethal toxin fragment described in the invention.

Sequence 546 AA:
XX
SO

Query Match	53.5%	Score 2780;	DB 20;	Length 546;
Best Local Similarity	99.1%;	Pred. No. 2.5e-143;		
Matches 541; Conservative	1;	Mismatches 4;	Indels 0;	Gaps 0;
QY	1	MNLVNKAOLQKMWYKFKRIQEDVEYVAILNALAEYHNMSSESVVEKYLKLDKIDNNLTNDYIL	60	
DB	1	MNLVNKAOLQKMWYKFKRIQEDVEYVAILNALAEYHNMSSESVVEKYLKLDKIDNNLTNDYIL	60	
QY	61	NTYKSGRKNALKKPKFXYLTMEVLELKNLSLTPPEKNLHFIWIGGOINDTAINYINQWKD	120	
DB	61	NTYKSGRKNALKKPKFXYLTMEVLELKNLSLTPPEKNLHFIWIGGOINDTAINYINQWKD	120	
QY	121	VNSDYTVKVFVDSNAFLINTLTKTIVESATNTLTSEFRENLTNDPEFDYKNKFKRKMELIY	180	
DB	121	VNSDYTVKVFVDSNAFLINTLTKTIVESATNTLTSEFRENLTNDPEFDYKNKFKRKMELIY	180	
QY	181	DKQKHFDIYKSIQISENPEFIIDNIITKYLSNEYSKDLKALNKYIEESLKNKITANNNDI	240	
DB	181	DKQKHFDIYKSIQISENPEFIIDNIITKYLSNEYSKDLKALNKYIEESLKNKITANNNDI	240	
QY	241	RNLKFADEDLVRLYNQBELVERWNLAASDIILIRISMLKEDGGVYLDVDLPGIQDPLPKS	300	
DB	241	RNLKFADEDLVRLYNQBELVERWNLAASDIILIRISMLKEDGGVYLDVDLPGIQDPLPKS	300	
QY	301	INKPDSITNTSWEMIKLEAIMKYKEYIPCYTSSKNFMDLDEEVQSPESALSCKSKSEIF	360	
DB	301	INKPDSITNTSWEMIKLEAIMKYKEYIPCYTSSKNFMDLDEEVQSPESALSCKSKSEIF	360	
QY	361	LPLDDIKVSPLEVKIATFANNNSVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE	420	
DB	361	LPLDDIKVSPLEVKIATFANNNSVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE	420	
QY	421	GTDFNTTWKIPSDKLASTSNEDNNMFMKIITNYLKVGFPAPDVIRSTINLSGPGVYTGAYOD	480	
DB	421	GTDFNTTWKIPSDKLASTSNEDNNMFMKIITNYLKVGFPAPDVIRSTINLSGPGVYTGAYOD	480	
QY	481	LLMFNDNSTNTHLLEPELRNPEFPKTKISQLTEQEITSLWSFNQARAKSQPFEYKKGYPE	540	
DB	481	LLMFNDNSTNTHLLEPELRNPEFPKTKISQLTEQEIRSLWSFNQARAKSQPFEYKKGYPE	540	
QY	541	GALGED 546		
DB	541	GALGED 546		

RESULT 6
AAR07503
ID AAR07503 standard; protein; 1254 AA.
XX
XX
AC AAR07503;
XX
XX
DT 25-MAR-2003 (updated)
DT 06-FEB-1991 (first entry)
XX
XX
DE Merozoite apical-end-localised protein (MAEP) insert 5.3.
XX
XX
KW Malaria; vaccine; Duffy blood group.

XX OS Plasmodium vivax.
XX PN WO9011772-A.
XX PD 18-OCT-1990.
XX PF 03-APR-1990; 90MO-US01849.
XX PR 06-APR-1989; 89US-0334270.
XX PS 05-APR-1989; 89US-0334041.
XX PA (UYNV) UNIV NEW YORK STATE.
XX PI Barnwell JW, Galinski MR, Wertheimer SP;
XX DR WPI; 1990-334616/44.
XX DR N-PSDB; AAQ06328.
XX PT Malarial apical end merozoite proteins and peptide(s) - used for
XX PT developing cpds. for treating, preventing and diagnosing malarial
XX PT infection
XX PS Claim 2; Fig 1a; 66pp; English.
XX CC A MAPP compound having a binding affinity for a Duffy blood group
XX CC antigen of primate red blood cells, is antigenic for the complete
XX CC protein, and may be used in diagnosis, treatment and vaccination
XX CC against invasion by P.vivax and P.knowlesi.
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 1254 AA;

Query Match 5.7%; Score 293.5; DB 11; Length 1254;
Best Local Similarity 19.7%; Pred. No. 1.4e-07;
Matches 207; Conservative 194; Mismatches 405; Indels 247; Gaps 45;

QY 119 KOVNSDYTKVFDVSNAPLNTLTKTIVSATNNTLESFRENLDPEFYKPY-----172
DB 7 KOTSFDEKKKSIKAEKMGNTLTK-----ELEKMDDEKNIEKEVEEAQIQYKRFIDHDVN 62

QY 173 -----KRRMIIYDKQHFIDYKQIEENPEFIIDNIITKYLNSYKSDLEALNKYIEE 227
DB 63 LMNDEVEKSLVMEK-----IELYKKEIDELKQ-----KT---NEY-KQDTSNFYTE 107

QY 228 SLNKTANNGNDIRNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDV 287
DB 108 QYNSATQSKA-----KIEQFT-----NIATTK-----KGTSDTSQDI 139

QY 288 DILPGIOPDLFKSINKPDSITNTSWEM-----IKLEAMKYKEYIPGYTSKN-----334
DB 140 NELESKEEVKRLQLVKQESNWEEMRKQILSMKOLLILNSETIAKEISNTQNALGF 199

QY 335 -----FDMLEDEQVRSFESALSSKSDKSEIFPLPDDIKVSPLEVKIAFANNVINOQA- 386
DB 200 RENAKTKLNTKTDLLQORVAAMIEEAKAKNNIDIALEDAQIDTEVSKIEQINREINWKD 259

QY 387 -----LISLKD-----SYCSDLVINOIKRYKI-----LNDNLNPSI 418
DB 260 EIKSYLSEIKYKDKCTTEISNKRKQDKIEFLEKFKPNEESNSKNVINEINENIRNSE 319

QY 419 NEGTDF-----NTWKIFSDKLASTN---ENMMFMKITNYLVKVGAPDVRSTINLS 469
DB 320 QYKLDIEADKQASTKVELFKHETTINIFKRESEILGVETKSKQKINKRAEDIMKEIERH 379

QY 470 GPGVYT--GAYQDLMFKONSTNHLLEPELRNFEFPKTKISQLTQEITSLWSFNQARA 527
DB 380 NSIEIQTVKGFQENLNKLNPNHYDNAEDLNKDKSTNAKV--LIETNLESV-----KHN 432

QY 528 KSQFEYKGYFGALGE-----DNDLPQNTVLDKDYVSKKILSSMKTRNKEYIHIVQL 584
DB 433 LSEITNIKQ-----GEKIYSKAKOIMQKATSENTAEKTLKVKDDQSQNVNYL----483

QY 585 QGDKISYEASCNLPSKDPYSSI-LYQKNIEGS--ETAYVVYVADAE-IKEIDKYRIPYQI 640
DB 484 --NQITERNLIVTEKRLNGIDSTITNIEGALKESKGYEIGFLEKIEEIGNRKLKVD 541

QY 641 SNKRNIKLTFIGHGKSEFNT-----DTFANLD-----VDSLSSRIETILNLAKADI---686
DB 542 ITKKSINST-VGNFSSLFNNFDLQYDFNKNINDYENKMGIEYNEFEGLSKISENLRNA 600

QY 687 -----SPKYI-----EINLIG-----CMFYSYSIYAETYPGKLLKIKDRVSELM 727
DB 601 SENTSDYNSAKTLRLLEAQKEKYNLLNKEEANKYLRDVKVESP--RFTFNNKESLIDKIN 658

QY 728 PSISODSITVSANQVEVRINEBGRKREILDHSGKWINKESIIKIDISSKEYISFNPKENKI 787
DB 659 EMIKKEQLTVNEGHGVQVQVQVNIKELVDE-----NLSLIDLKQATGKI-----EEIOK 707

QY 788 IVKSKYLHELSTLQEIERNNANSSDIDLEKKVMLTE-----CEINVASNIDRQIVEGRIB 843
DB 708 ITHSTLKNKAKTILGHVDTSAKYVGKITPELALTTELLGDACLKTAQELKPFESKNVVL 767

QY 844 AKNLTSDSINYIKNEFKLIESISD-----SLYDLKHQGLDDSHFISFEDISK-TENG----895
DB 768 TENMSKNT-----NELDVHKNIQDAYKVALEILAHSEIDITKQ-----KOSSLIENGNOI 818

QY 896 -FRIRPINKETGNSIFIEKEIFSEYATHISK---EISNI--KDTIFDNVANGKLVKYN 949
DB 819 YLKVVLINQYKNKISSIKSEAEVSVKTVGNVSKHSELSKITCSDKSYDNI-----IA 871

QY 950 LDAHEVNTLSNAFFIQ-----SLIEYNTKESLGNLSVAMK-----986
DB 872 LEKQTELQNLNRNFTQKTNNTNSDSKLEKIKTDPSLKNALKTLEGEVNAKASSDNHEH 931

QY 987 VQVYAQLPSTGLNTITDASKVVELYSTALDETI 1019
DB 932 VQSKSEPNPALSEIEKEETDIDSNTALDELL 964

RESULT 7
AAW24575
ID AAW24575 standard; Protein; 1254 AA.
XX AC AAW24575;
XX DT 25-MAR-2003 (updated)
XX DT 10-NOV-1997 (first entry)
XX DE Merozoite apical-end protein clone 5.3.
XX KW Merozoite apical-end protein; MAEP; Plasmodium vivax; antigen; malaria;
XX KW human; erythrocyte; antimalarial vaccine; antibody; blood stage parasite;
XX KW Duffy blood group antigen; red blood cell; therapy.
XX OS Plasmodium vivax.
XX US US646247-A.
XX PD 08-JUL-1997.
XX PF 04-OCT-1991; 91US-0792865.
XX PR 04-OCT-1991; 91US-0792865.
XX PR 05-APR-1989; 89US-0334041.
XX PR 06-APR-1989; 89US-0334270.
XX PR 03-APR-1990; 90MO-US01849.
XX PR 02-NOV-1990; 90US-0608639.
XX PA (UYNV) UNIV NEW YORK STATE.
XX PI Barnwell JW, Galinski MR;
XX DR WPI; 1997-362995/33.
XX DR N-PSDB; AAT80072.
XX XX

CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.

XX Sequence 1516 AA;

Query Match 5.4%; Score 280; DB 21; Length 1516;
Best Local Similarity 19.2%; Pred. No. 9.6e-07;
Matches 230; Conservative 187; Mismatches 351; Indels 432; Gaps 57;

QY 6 KAQLQKVVYKRIQDEYVAIILALEEYHNKSSVVEKYLKLDINLTD-----NYL 60
DB 131 KTAELKLVNYQO-----RTLNSMKSKKNKNDNS-----NNIEDKTNTPNKT 173
QY 61 NTYKSG-----RNKALKKFKYL-----TWELVLE 85
DB 174 NTQKSNQNTKNTPNKINADISKSLIQIYDDIKEKSLSLVHGVPVSKVDLT 233
QY 86 LKNSLTPVEKNLHFI-----WIGQINDT-AINYINQKDVNSDYTVKVFYDSN-AF 136
DB 234 ICNDLSKIKKIFMITDFGPVLFGEQDGMGTVENINKLDNRNKD-----ENNL5Y 286
QY 137 LINTLKTIVESATNTLSFRENLDNPFYKFKRMEIYDKQHFIDYYSQIEE 196
DB 287 SINYNKVQDVNNNDKDKKENINEVRD-----QKNYV-YKNK----- 326
QY 197 NPEFIIDNIKTVLSNEYSKLEALNKYIBESLKITANGNDIRNLEKFADELVLRL-- 254
DB 327 -----ENINNYLDDDEK-----EDQNKGVYNNDDI-----DEQIRKQH 364
QY 255 MARKKYIESIPKTFGLCWRPVDIIDISNTYEMLE-----ISETLKVHENKFKQHL 418
DB 365 MARKKYIESIPKTFGLCWRPVDIIDISNTYEMLE-----ISETLKVHENKFKQHL 418
QY 282 GYVLVDVILPGQDPLFKINPKDSITNTSWEMIKLEAMKYKEYIPGYTSKQFMDLEE 341
DB 419 NVLDENNSTFVNMNLLKNIN-----YK-----KNDLLIEGG 450
QY 342 VORSF-----BSALSSKSKDSE-----IFPLDDIKVSPLEVKIAPANNVSN 384
DB 451 EKKSFLINLVDSVSSNSRLDENIERGKINMFTNDEKSIINN-----YNNNNNN 506
QY 385 QALISLSDSCYSDLVINOIKRYKILNDNLNPSINEGTFDNTMKIFSDKL--ASISND 442
DB 507 NNNNNNDNNNDNDVIEHNKNNNIYDNKYN-----VECSSEKINDNGISKN 554
QY 443 NMMFMKITYNLKVGAPDVRSTINISGPGVTGAYQD-----LMPKONSTNIHLLEPE 497
DB 555 --INILEPNNLDT-----SNIFLEGKDEYKVYVYVKEIRIPLPKEINKEI----- 599
QY 498 LRNFEPKTKISQLTQEITSLW-----SNOARAKSQFEEYKGY--- 538
DB 600 ----FEKLPLKQVQIILQDIKEEYNTNRJKAISKDDMDVFSQVQETVYRMKIDTFEIE 656
QY 539 -----FEGALGDDDLDFQAQNT-----VLDKDYVSKKILSSMKTRKEYIHYIV 582
DB 657 KLKIKMAENIQSVEGEL--LINKOLSKNTDNNIKDYNVLQKK---KSKKKKKFLNDIL 711
QY 583 QLOQKGISVEACNLPFSKDPYSILYQKNIEGSETAYYYYADAETIKEDKIRYPQISN 642
DB 712 N-----TYNFTTESKYQDL-----YVKGEEKSKDIKNQIDPVTQE 746

QY 643 -KRNIKLTFIGHKSEFNTDTFANLDVDSLSSEIETILNLAADISPKYIEINLLGCNMF 701
DB 747 CYRNDIIRDTHDKS-----DIFKNIKDNNKKYIYNLEOEINEKK-----NYN 794
QY 702 SVSYVAETYPKLLKLDKRV-SELMPSISQ---DSITVSANQYEVRI-----NEEGKRE 753
DB 795 KNNDSNKTFF-----FLKIENEFKOLLDDSDQIFGDSLADIKEYNYTADNLDNNENKS 850
QY 754 ILDHSGKWKNEESIIDISKE---YISFNPKNK--IIVSKYLHELSTLLQEIENNA 808
DB 851 LYEDGENFTIRNEPTNTEYEKNNIYISDEQYNEEDIFFKDK-----IKEKKNND 903
QY 809 NSSDIDLEKKVMLTCEINVASNIDROIVEGRIBEAQNLTSD-----SINYIKNEFKLIES 864
DB 904 TSSD-DFB-----NCSVQ-----EKIYVNEKIEYNNKNDKSSSSSIILIEIKYKE 951
QY 865 ISDSLY-----DLKHQGLDDSHF----- 883
DB 952 KXDELVSPNLCVLLDDEFHSDNLENNYISVSSDDMKTNVSKNNITGVKENKVDKTNVEYD 1011
QY 884 -----ISPEDISKTEGFRIRFINKETGNSIF---IETEKEIFSEVATHISK--- 927
DB 1012 KKGDDGVIEISFEDSHKLEES---KF---DNNNIYDNDDELEKLSKDYISDQDKQHN 1065
QY 928 EISNIKDTIFDNVNGKLVKVNLDAAHEVNTLNSAFFI-----QSLIETVNTTSESLSNLSV 983
DB 1066 NIYNIERGEDBRENEFVENKIQSTESHKSNE-----FICTENKSLRKQYMSKEDISNVR 1120

RESULT 10

AAB18294

ID AAB18294 standard; Protein; 980 AA.

XX AC AAB18294;

XX DT 07-NOV-2000 (first entry)

XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:152.

XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;

XX KW antimalarial; malaria; protozoaside; infection; insecticide.

XX OS Plasmodium falciparum.

XX PN WO200025728-A2.

XX PD 11-MAY-2000.

XX PF 05-NOV-1999; 99WO-US26796.

XX PR 05-NOV-1998; 98US-0107131.

XX PA (HOFF/) HOFFMAN S.

XX PA (CARU/) CARUCCI D.

XX PA (GARD/) GARDNER M.

XX PA (VENT/) VENTER J C.

XX PI Hoffman S, Carucci D, Gardner M, Venter JC;

XX DR WPI; 2000-365347/31.

XX PT Proteins encoded by chromosome 2 of the human malarial parasite,

XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the

XX PT diagnosis of P.falciparum infection -

XX PS Disclosure; Page 362-365; 577pp; English.

XX CC The present invention describes proteins and their fragments (I) encoded

XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.

XX CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)

XX CC vaccines against P. falciparum infection comprising (i) or (ii).

XX CC (i) and (ii) are useful for the development of vaccines against

P. falciparum infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with *P. falciparum*. Furthermore, (I) (especially when they are rfiins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the Plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

AA	Sequence	980 AA;
AA	Sequence	980 AA;

Query Match	5.2%	Score 272;	DB 21;	Length 980;
Best Local Similarity	20.3%	Pred. No. 1.5e-06;		
Matches 226;	Conservative 180;	Mismatches 393;		
		Indels 316;		Gaps 52;

Qy	6	KAQLQKQWY-VKFRIQDEYVAILNALBEYHNMSSESVVEKYKLKLDKINNLTNDVNLTYK	64
Db	39	KEGKEGIYELKEKLNDDE-KIINDLKKNYS-----QVYKMDQYKEKNNLINEIN	89
Qy	65	KSGRNKALKKFKEYLTMEV---LEKKNLSLTPV-EKNLHFIIWIGQINDTAINTINOMKD	120
Db	90	K-----LKUFIEBNKMTVERGNEWNKKLEEMQKQKELINNLNDSDELKNCIEQVNS	143
Qy	121	VNSDYTVKFFVYDSNAFLN-----TLKKTI-----VES-----ATNNTLESFR	158
Db	144	VSRN--MANVEKEKENIINELQILRMKNDITMRKISKFVEQEKVLFKFLYLLNNDIFSKN	201
Qy	159	ENLNDPEFDYNNFVYKBEWELIYDKOKHFD-----YYSQIENPEFIIDNIKT---	208
Db	202	EKUND-----MOKLNDVNEKYKNIVECLNNKYTEHKQIEKKIERI--NTLKQNY	251
Qy	209	YLSNEYSKOLEALNKYIE-----ESLNKITANNNGNDIRNLEKFADELDV	252
Db	252	YLKEYDLKNEKEKNEHGGKLEHLSHCYEEQNKLNEEBIKRRNSFIKNKDR-KIDLL	309
Qy	253	RLYNQELVERWNLAASDILRISMLEDGGVYLVDLILPGLOPDLFKSINKPDSITNTSW	312
Db	310	TNIENELLKK-----KEINNIKMEQONVTKNNEQLLKQIDENEKKNHEVHNKLQN---	360
Qy	313	EMIKLEAIMKYKEYIPGYTSKNFMDLDEEVORSFESALSSKSDKSEIFLPLDDIKKVSPL	372
Db	361	ELIKRE-----LONKCSKDIJFCCKEKEDK-----TKNLEDDLULE	396
Qy	373	VKIAFANNSVINOALISLKSDYSCLDVINOIKNRYKILNDNLNPFINSINEGTDFNTMTKIFS	432
Db	397	KKKCIEN--LKDELIINIK-----KKMEDKMHMTNE-MDLLSNKVEELNRINKTYE	443
Qy	433	DKLASISNE-DNMFMKIKITNLYKVGFPADPVRSITNLSGPGVY-TGAYQ-----DLL	482
Db	444	KNIVELANNELDVIKKKLNDEEFLK---EEEEKKKIDM---VYKIKYEIOIKIENKENEID	496
Qy	483	MPKDNSTNIHLLEPRLNFPFPKTKISQOLTQEQITSLTSWFSNOARAKSQSFEEYKKGYPEGA	542
Db	497	SLKKQEQNLHLVKNE-----ELNEKEIILKNKYDK-EINMLIEQYNKKI-----	539
Qy	543	LGEDDNLDFQAQNTVLDKDYVSKKILSSMTKRNKEYIHYI-----VOLQGDKISVYASCNL	597
Db	540	---QEEQDMLNKKIKSMDDQTHKNQIIEEMQEBNKKELKELKNVCDMLNLSQILIKENEXHM	596
Qy	598	FSK--DPYSSILYQKNIEGSETAYYYYVADAIEKB-IDKYRIPYQIJSNK-----	643
Db	597	QEKVEEYANLKKQK-----DQELKNIIOEYDERIEIQNEMEDIVNDCEBK	642
Qy	644	-----RNIKLFTFIHGKSEFNTDTFANLDVDSLSSESTIETILNLAKADISPKYIEINLLG	697

643	LKQAKINKKLT-----TATNMANNMWMNDENLKEKDKKINDLMKOMEKKKEEINKLV	696
698	-----CNMFYSIVAEETYP-----GKLLIKIKDRVSELMPSISQDSITVSANQ	741
697	EEKSKLEHSHVKIQNEMSLLVQONEKLEEMGLSRIAIO-MEEIKOMKEVEEKKKNE	755
742	YEVRINEGKREILDHSGKWINKEESIIKDISSKEYISFNPKNENKIIIVKSKYVHELSTLL	801
756	EERKKNBERKKNBERKK--NEEEKKNBERKK-----NEEEKKLEKDKHQFE-----	804
802	QEIRNNANSSDIDLEKKVMLTECEINVASINDROIQVEGRIEBAKNLTSDSINYIKNEFKL	861
805	-----EKKERM-----EYEHQKEDR-----KRDKKKKGHSSDKEEKYNNKEKT	844
862	IESISDSYDLKHQGLDSDHSPISPEDISKTENGFRIRFINKETGNSIETETEKEIFSEY	921
845	KEKSSNLLP-----DEEYIIQLEEL-----RDTGENCFI-----Y	874
922	ATHISKEISNKDTIFDNVNGKLVKKVNLDAAEVNTL-----N	960
875	LKLSLSEL-----DVIINKLKSDDALLN-DAPNKINLAITSWNIPNEENKEGDNITTVEN	929
961	SAFFTQSILIENTTKESLSNLV--AMKVQVYAQL	993
930	TATEGNITIDENTTEVENMNEEVYKIFSVKEYDML	964
RESULT 11		
AAAB18241		
ID	AAAB18241 standard; Protein; 1121 AA.	
AC	AAAB18241;	
DT	07-NOV-2000 (first entry)	
XX		
DE	Plasmodium falciparum chromosome 2 related protein SEQ ID NO:98.	
KW	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;	
KW	antimalarial; malaria; protozoasidae; infection; insecticide.	
OS	Plasmodium falciparum.	
XX		
PN	WO200025728-A2.	
XX		
PD	11-MAY-2000.	
XX		
PF	05-NOV-1999; 99WO-US26796.	
XX		
PR	05-NOV-1998; 98US-0107131.	
XX		
PA	(HOFF/) HOFFMAN S.	
PA	(CARU/) CARUCCI D.	
PA	(GARD/) GARDNER M.	
PA	(VENT/) VENTER J C.	
XX		
PI	Hoffman S, Carucci D, Gardner M, Venter JC;	
XX		
DR	WPI; 2000-365347/31.	
XX		
PT	Proteins encoded by chromosome 2 of the human malarial parasite,	
PT	Plasmodium falciparum, useful as antimalarial vaccines and in the	
PT	diagnosis of P.falciparum infection -	
XX		
PS	Disclosure; Page 229-232; 577pp; English.	
XX		
CC	The present invention describes proteins and their fragments (I) encoded	
CC	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.	
CC	Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)	
CC	vaccines against P. falciparum infection comprising (i) or (ii).	
CC	(i) and (ii) are useful for the development of vaccines against	
CC	P. falciparum infection. (i) and polyclonal antisera or a monoclonal	
CC	antibody raised to immunogens comprising the sequences of (i), are	
CC	useful in the detection of infection with P. falciparum. Furthermore,	

Disclosure: Page 229-232: 577pp: English:

The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, *Plasmodium falciparum*. Also described are: (1) nucleotide sequences (II) encoding (I); and (2) vaccines against *P. falciparum* infection comprising (I) or (II). (I) and (II) are useful for the development of vaccines against *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with *P. falciparum*. Furthermore,

(I) (especially when they are refined or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAY70078 to AAY70287 and AAB18144 to AAB18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

Sequence 1121 AA;
xx
SQ

Query Match	5.1%	Score 266.5;	DB 21;	Length 1121;
Best Local Similarity	20.2%	Pred. No. 3.5e-06;		
Matches 238;	Conservative 183;	Mismatches 372;	Indels 385;	Gaps 55

Qy	5	NKAQLOKQWVYKFRIQ--DEYVAILNALEEYHNMSE-----SSVBEKYKLKLDI 52
Db	41	NEKRTNKIKYKSKAQSLFDKGLNIHDKLILFKNLPKYKCAKECISAKEVYKL-----95
Qy	53	NNLTDNYLTNYKSGRNKAL-----KKFKEY-LTMEVLELKNNSLTPVEK- 96
Db	96	-----LDEYKCKPCNYISLCOIIQSVKIFDBELDKTFTDYNFVIEVKNIDKQVLANKINEI 148
Qy	97	-----NLHFITWIGQINDTAINYINO-----W-KDVN-----SDYTVK 128
Db	149	YPKNKDITFHRREILGKICNKIMSYIHEMNGNELIHFLYPRFWNKDQKLLIFNYNYFN 208
Qy	129	FVYDSNAFLINTLAKTIV--ESATNNTLESFRENLDPEFDYKRFYKRMIEIYIDKQKH 185
Db	209	YVFDHMYLFNHEIYKLLFIPNKYLNNSNIPFNKLIQ-EMEFNLYYFREIK---NECN 263
Qy	186	FI-----DYK--SQTIEHPEFIIDNIIKTYLSNEYSKOLAEALNKYIEESUNKITANN 236
Db	264	YIIKMKKKEIYKCKPAKPHENVDH-IDN-----EKILNILRLYVDNSILDIDINN 312
Qy	237	GNDIRNLEKPADEDLVRLYNQBELVERWNLAAASDILRI--SMLKEDGGVYLDVDILPGTO 294
Db	313	-----KMLCNLNNLINE--NIEYISKLLNFYCTLIKK--GKY-----346
Qy	295	PDLFKSINKPDSITWTSHEMIKLEAIMKYKEYIPGVTSKNPDMLEDEEVQBSFESALSSKS 354
Db	347	-----DNDMTYIKLKEVIAKATHILCDKTKNLETFCSDI--DYSTLLNSLN 390
Qy	355	DKSEIFLPLDDIKVSPLE-----VKIAFANNSVINOALISLSDVSCYSLVINOIKQRYK 408
Db	391	NKPIKLIIDKQKILFYECLLKILNIKFPVNFQSCISLISLKNY-----YNILRNVY 445
Qy	409	ILNDLNPSINEGTDFNTMK-----IFSDKLASISNEDNMFMK--ITNYLKVGFA 459
Db	446	IVNVVL-----FNDIMFSLYLCNIFLGKRIKTENENAVLIHNNDOQNYSN----492
Qy	460	PDVRSNTINSLGPGVVTGAYQDLMPKONSNTNHLLEPELRNPEFPKTKISQLTBEITS 519
Db	493	-----KENIKOI-IIQKRIKEYIFYK-----512
Qy	520	WSPNOARAKSQEEYKGYFEGALGEDDNLDPAQNTVLDKDYVSKILLSSMKTR-NKEYI 578
Db	513	-----MENYKDPHFK-----LKSDLSLISLKLNTFPVKINEVYN 546
Qy	579	HVIVLOQGDKISYEASCNLFS-----KDPVSYSL-----YOKNIEGETAY 619
Db	547	SYDFVLLFNNI-----SCLINFLNRSVKKYDYIYIINDLSFYVYKIKNDRTKKKK 602
Qy	620	YYYVADAEIKEDKYRIPVOISNKNRKNILKTFIGHGKSEFNTD-----TFAN--LD 667
Db	603	NFFLLSSSMKELICKNI-LSVGNR-----YIKHLEEDNFOKDQYVCSLFLNNLFPD 655
Qy	668	-----VDSLSSEIETILNKAAD--ISPKYIEINLLGCNNMFSYSI--YABETYPGKLL 716

Db	656	KIIHFYIYNLWCHVYKTYNYFKCNKLNEDIISLLLTCSKFPQFIENNSMDRYCRKEL	715
Qy	717	LKIKDRV-----SELMPISQDSITVSANQYEVKRIINEEGKREIILDHSGKWIN	763
Db	716	IHLKNIIDDLKYNLYTKISIDNISKIFISLSNPKYTCVENLLESLSQSEFKVT	775
Qy	764	K-----EESIIDKISKE-----YISFPNPKENKIIKSKYLHELSTLLQEIIRNA	808
Db	776	KTSKKGGIHMNDNLLDNNSNCEKYEHRYYEK-KENLFINLNKIIIECLIKLNIPLYLKK	834
Qy	809	NSSDIDLEKKVMLTECEINVASNIDRQI-----VEGREEEAKNLTSS	849
Db	835	KKTLYLYKQSL---CPINLKENILKILKILYIANNLYMYGYVCEMLERVLSSHKEQNL	891
Qy	850	DSINVIKN-EFKLIESISDSLYDLKHQGLDSDSHFISPEDISKTEGFRIRPINKETGNS	908
Db	892	FSYNNKVNVEHKMFDKI-----LCH-----ISEDDYIEMSNMIVV-----	926
Qy	909	IFITEKEIFSEYATHISKE-----ISNLIKOTIFDNVNGKLKVKVNLDAAEHVNTLNS	961
Db	927	LFYDYLKININERQSNILRNNSNDRFIDEIKEKKYKLNNTLIKHNNVKNLYEKS-N	985
Qy	962	APFTQSLE-----YNTTKESLSNLSVAMKQVYVQ	992
Db	986	NGNISNLKDDKNKHNHNVEMDLIDNKNENKKIOEKQ	1023
RESULT 12			
AAE00425			
ID	AAE00425	standard; Protein; 2184 AA.	
AC	AAE00425;		
DT	19-JUN-2001	(first entry)	
XX			
DE		P. falciparum telomerase reverse transcriptase full length protein.	
KW		Telomerase reverse transcriptase; TERT; ever shorter telomere;	
KW		EST; therapy; stomach cancer; malaria; vaginal candidiasis.	
OS		Plasmodium falciparum.	
Key		Location/Qualifiers	
FT	Misc-difference 330	/label= Leu, Ile	
FT		/note= "Encoded by WTA"	
FT	Misc-difference 335		
FT		/label= Asp, Gly	
FT		/note= "Encoded by GWT"	
PN	WO200127287-A2.		
XX			
PD	19-APR-2001.		
XX			
PF	10-OCT-2000; 2000WO-US27825.		
XX			
PR	13-OCT-1999; 99US-0417485.		
XX			
PA	(RERE-) RES & DEV INST INC.		
XX			
PI	Long DM, Metz AM, Love R;		
XX			
DR	WPI; 2001-266411/27.		
DR	N-PSDB; AAD03729.		
XX			
PT	Telomerase reverse transcriptase genes and proteins from Plasmodium		
PT	falciparum and Candida albicans are used to detect infections of these		
PT	pathogens in mammals -		
XX			
PS	Claim 10; Page 104-111; 138pp; English.		
XX			
CC	The invention relates to identification and use of telomerase		

XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) especially when they are rifins or secreted or membrane proteins
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.

XX Sequence 1712 AA;

Query Match 5.0%; Score 258.5; DB 21; Length 1712;

Best Local Similarity 20.3%; Pred. No. 1.7e-05;
Matches 219; Conservative 163; Mismatches 394; Indels 301; Gaps 52;

QY 34 YHNMSESSVVEKYLKDOI--NNLTDNYLNTYKSGRNKALKFKPEYLTMEVLEKNLSL 91
DB 48 YNN-----VDKIPNKEIYNNIKSNNIQVRVQNQNNNEEKKKEANYT-----CVNNKY 96
QY 92 TPVERKULHIWGGQINDTAINYI-----NQWQVNSDYTVFVYDSNAP 136
DB 97 VTLKNKVH---VNKYVNSNINKITVPIIKCSNYKIKNPIKSHLSNENKFKVLSN-- 151
QY 137 LINTLKK--TIVESATNNTLEGFR-ENLNDPE-----PDYKNF---YRKRMEIYDKQK 184
DB 152 -PSNINKGSHKNDVINETMDQHSQOLANDNIIKLLYDYCFIPREDTIKTNTNISYNKN 210
QY 185 HFIDYKSGIENPEFIIDNIITYLSN--EYSKOLEALNKYIEESLNKITANNNDIRN 242
DB 211 SPKDN-----EENINTMDNNIKSNSSSYCSYS-----NKNQNNVNHHTL 251
QY 243 LKFAEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVLDVDILPGIQDLPKFSIN 302
DB 252 KTEFLNEKNSHTQNEQ-----SIFLLDGLQ-----NNHN 280
QY 303 KPDSINTSGWEMIKLEAIMKYEYIPGYTSKNFD---MLDEEVORSFESALSSKSDKE 358
DB 281 SATKFNNIYD--NNSLVNYK-----SDKGIDLHNKMKIETDKNGIITLEKKKHDEK 332
QY 359 IFPLPDDIKVPLEVKIAFANNNSVINOALISLKSQY---CSDLV-----INQIKN--R 406
DB 333 YI---NNIFLPLNDN---SNNVITTC--DNKESYRNSTSDMINKIFEKMWNEKKNILK 384
QY 407 KYILNDNLNPSINEGTD--FNTTMKIFPSDKLASISNEDNMFMKIITNLYKVGFPADVRS 464
DB 385 MKNFNDVIKKKITMAKEKILNSSTNNMKKVSFYNKSKDEDLFNEKENSY-KYGVKRENOE 443
QY 465 TI-----NLSGPGVYTVGADLLMPKDNST--NIHLEPELRNFEPP----- 504
DB 444 DINVIKNKRNININDNDNINIIKNDVSKRNIIHNKKKRRDDPPFNNSAGLLDFFDL 503
QY 505 -RTKISQLTEQEITSL-----WSFNQARAKSQFEYKKGYPFEGALGED----- 546
DB 504 CRRKVLIELKNVQSSKKKKNKILTNHNSSDNQCHSSDNQCHSSDNQCHS 563
QY 547 -DNLDPAQTVLDKQVYSKLLSSMTRNKEYIHYIVQLQGDKIYSYASCNLFPSKDPYS 605
DB 564 SDNQNCDSNACNKKDEKRRKKKKIKKKNM-----KNKSNKSNK--KRETKSK 611

QY 606 ILYQKNIEGSEYATYYVADAEI--KEIDKYRIPYQISNKRNIKLTFIGHGKSEFNTDTF 663
DB 612 KISNNNNNDNMNQDNDMGQRINNENMDQNV--NIQNEGN-----GFNNNNKNNDDL 663
QY 664 -----ANLVDLSLSSEIETILMLAKADISPKYIEINLLGCNMFPSYIYABETYPGKLLLK 718
DB 664 NYIYSPNMINHLSSTCE-----KKNKEDNKNMDNKPNS-----SSKMK 703
QY 719 IKD---RVSELMPISODSITVSANQVEVRINEGKREILDHSGKWINKESIIKD--I 772
DB 704 IPEITNNSNEKIVNVDNDEMUVYHNLTVLVNKBQ-----GVTEESSICKRTYF 753
QY 773 SSKEYISFPNKENIIVKSKYLHEISLTLLQEIINNANSSDIIDLEKKVWLTE----- 823
DB 754 VQFYDYSVMRNEKITDDNMQVEDIYVKNENIKRTLKGDGHDVKTNNMLSEDNSYASGLW 813
QY 824 -CEINVASH-----IDRQIVEGRIEAKNLTSDSINYIKN----- 857
DB 814 GNEINFISNNENCLNSYDISCDERKIPNBEEOBELCSNNI-LVKDIEBKQMGCLFPFE 872
QY 858 --EFKLIESISDSLYDLKHQGLDDSH--FISFEDI---SKTENGFRIRFINKETGNS 908
DB 873 ICVFRINEKNEHGHENLRKNNDTHKMYSSYEMIQNINKOSTNPFCCKDEMEKSGQTN 932
QY 909 IFIE-----TEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAHEVN 957
DB 933 LFYDNYINSVDITKLELNKNCYQHINYEQN-----LIKKENSYAA-ENN 976
RESULT 14
ABB54799
ID ABB54799 standard; Protein; 1090 AA.
AC ABB54799;
XX
DT 16-MAY-2002 (first entry)
XX
DE Lactococcus lactis protein yoiA.
XX
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS Lactococcus lactis IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PF 11-APR-2000; 2000FR-0004630.
XX
PR 11-APR-2000; 2000FR-0004630.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
DR WPI; 2002-043418/06.
XX
PT New nucleotide sequence useful in the identification or Lactococcus
XX lactic acid and related species -
XX
PS Claim 6; SEQ ID No 1501; 2504pp; French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX	Sequence	1090 AA;	Query Match	4.9%; Score 256; DB 23; Length 1090;	
SQ			Best Local Similarity	19.5%; Pred. No. 1.3e-05;	
			Matches 226; Conservative 185; Mismatches 370; Indels 378; Gaps 59;		
QY	11	KMYVVFRIQEDRY----	VAINALAEVHN--MSESSVVEKY-LKLKDINNTDYNLTNY	63	
DB	91	KKAYLKLNENLDFNLLSOKIADKQDYNSEMIKYEEVKKYQESLUKEYQETKEYENVI	150		
QY	64	KKSGRNKALKKFKYELTMEYLELKNNSLTPVPEKNLHFIWIGQINDTAINYNQWKDVS	123		
DB	151	--NDRKAAQKPEE-----NSLALKN-----INDEVQNLIDK-KSVKN	185		
QY	124	DYTVKVFYDSNAPLINTLTKTIVESATNTNLTESFRENLDNPDY----	NKFRKRMEIYY	180	
DB	186	D-----AF-IBEKELAVEQEFAAES---QLNGIVSVYELLKKNYQEKKATFE	231		
QY	181	DKQKHFDIVYKSOIEENPEFIIDNIITKTVLSNE-YSKDLEALNKYTEESLKNITANNND	239		
DB	232	ESRKY--RHANTIKDWEMLIKDFEKTNSKEKFEONLOAIDDKVKAEEEE--ANISNE	287		
QY	240	IRNLEKFADEDLV-----RLYNQELVERNLAFA--SDILRISMLKEDGGVYLD	286		
DB	288	IYN---FVPSLVAMTEKYEHKKVYEAAL-AYNISKGYDAEYAKFSELKEN-----	337		
QY	287	VDILPGIQDPLFKSINKPSITNTSWEMIKLEAIMKYEYIPGYTSKNFMDLDEEVQSF	346		
DB	338	-----FQEIIN---SLKNPSQTVI-----GY-----ITRDM	359		
QY	347	ESALSSKSKSEFLPLDDIKVSPLEVKIAPANN--SVINQALISLKSQVCSDLVINQIK	404		
DB	360	EAKNRVDNLAKDFPLEE-----KFNLOKNIFEQISTSYLVKN-----LVDEKK	405		
QY	405	NRKYILNDNLNPSINEGTFNMTMKIFSOKLAGISN---EDNMFMIKITNY--LKVGFA	459		
DB	406	GRSKVSGNTLTETLNAQKKTWLVLEVKSKYGLLNHORGQSEAFKVIQQVYEELEKEFQ	465		
QY	460	PQVRSINISGPGVYTGAYQDMLFMKDSTNIHLLPELRFNFPFTKISQLTE-----	513		
DB	466	KEQSFSFKIK-----EYTAPETRYNTK-----ESYDIAKEKLAQSEENLEIVK	510		
QY	514	---QELTSLSMFSNQARAKSFQYKKG--YFEGALGEDDNLDF-AON--TVLDDKDV---	561		
DB	511	KAYSDFIAGLRVATVAKESILEEYKSEKFNELTDTKEGFEFVQNNFTGLDETQKGE	570		
QY	562	-VSKKILSSMKTNRKNEYIHYI-----VOLQGDKIS--YEASCNLFSKDPYSSI	606		
DB	571	QLCLNLTEQAKTWHDDYPOYSHSYKTLIPYTKVNSQLEKLSQYEVWFNDYQNLNKSSE	630		
QY	607	L-----YQKNEGSETAYYYVADAETKE-----	630		
DB	631	IVRTOYQEQEQ-----RYNDAQELYNQVSEFBNHCLDLNRVVLKKDNEFETALNNLQ	683		
QY	631	-----IDKYRIPYQISNKRNIKLTFIGHKSEFN--TDTFANLD--	667		
DB	684	FIPKNTLTPPIGEATGSNFIDEFNIPETIEVSSSKLT-----NKFELSQWTTYNQVLA	739		
QY	668	-VDSLASEI-----ETILNLAKADISPKYIEINLGLCNMFYSIYA	707		
DB	740	LVDISIKELQALTEYQNLQSIFIKVGGKRVLDGNSDIAYR-----GTNI-----	785		
QY	708	EETYPKLLKIKDRVSELMPSISQDSITV-SANQY-----EVRINE-----	754		
DB	786	-ESYLERLAINFKWEQSEYVKKRQVFSLLVKNYRQYLEESVKLNENIARMSKNGLOEI	844		
QY	755	LD-----HSGKWINKK--BSIIKDISKEVISFNPKENIKVSKYHLHELSTL	800		
DB	845	INAINSLPNYMDIYSTSYRNVELLASGALGAEVSDLTLHR-NIDELKEINKVTFEHLAL	903		
QY	801	L-----QEIRNNANSSDIDLEKKVMLTECEINVASNIDRQIVGEIRB---EAKN	846		

DB	904	IDENLNTINHPSEIRSSYDQSIINLRKQ--LDMEKITGIPTRKVMIESSLPNDLIDLKS	961		
QY	847	LTSDSIN-YIKNEFKLIESISDSLYDLKHQGLDDSHFISPEDISKTENGFRIRFINKET	905		
DB	962	YFTGALNENVKOLFSTIISSSSFINSVKKVN-----QFKRVQNLLSV---NEG	1006		
QY	906	NSIETIETEKEIFSYATHISKEISNIKDTIPDNVNGKLKVKVNDLDAAEVNTLNSAF--	963		
DB	1007	GNAIILPEVPIAPDKKELEFSLVIATIK-----NIDAPVKVEKLESAYVK	1051		
QY	964	-----FIQSLIE	970		
DB	1052	PEKPENLADKPSYVEDIVE	1070		
RESULT	15				
ID	ABP73774				
AC	ABP73774;				
XX	30-JAN-2003	(first entry)			
DE	Candida albicans essential protein SEQ ID NO 7611.				
KW	Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;				
KW	signal transduction; DNA replication; cell division; growth;				
XX	proliferation; Candida albicans; fungicide; antifungal.				
OS	Candida albicans.				
PN	WO200253728-A2.				
PD	11-JUL-2002.				
XX	26-DEC-2001; 2001WO-US49486.				
XX	29-DEC-2000; 2000US-259128P.				
PR	20-FEB-2001; 2001US-0792024.				
PR	22-AUG-2001; 2001US-314050P.				
XX	(ELIT-) ELITRA PHARM INC.				
PI	Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;				
XX	WPI; 2002-566694/60.				
DR	N-PSDB; AB232324.				
XX	Constructing strains for identifying gene products as effective targets				
PT	for therapeutic intervention, by inactivating in the strain one allele				
PT	of a gene and placing other allele of the gene under conditional				
PT	expression				
PS	Claim 44; SEQ ID NO 7611; 167pp + Sequence Listing; English.				
CC	The invention relates to constructing (M1) a strain of diploid fungal				
CC	cells in which both alleles of a gene are modified, comprising modifying				
CC	one allele by insertion or replacement by a cassette having an				
CC	expressible selectable marker and modifying other allele by				
CC	recombination, of a promoter replacement fragment with a heterologous				
CC	promoter, so that expression of the second allele is regulated by the				
CC	promoter. (M1) is useful for constructing a strain of diploid fungal				
CC	cells in which both alleles of a gene are modified. The diploid fungal				
CC	cells having both alleles modified are useful for identifying a gene that				
CC	is essential to the survival or growth of a fungus, a gene that				
CC	contributes to the virulence and/or pathogenicity of a fungus, a gene				
CC	that contributes to the resistance of a diploid fungus to an antifungal				
CC	agent, an antifungal agent that inhibits the growth of a diploid fungus				
CC	and for identifying a therapeutic agent for treatment of a mammalian				
CC	disease. (M1) is useful for identifying a compound which modulates the				
CC	activity of a gene product, preferably enzymatic activity, carbon				
CC	compound catabolism, biosynthetic, transporter, transcriptional,				
CC	translational, signal transduction, DNA replication and cell division				

CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX

SQ Sequence 1948 AA;

Query Match 4.9%; Score 256; DB 23; Length 1948;
Best Local Similarity 19.3%; Pred. No. 2.7e-05;
Matches 233; Conservative 207; Mismatches 441; Indels 324; Gaps 53;

```
QY 4 VNKAQLQKVVYKFIQEDVEYVAILNALBEYHNMSESSVVEKYKLKQINNTIDVNLNTY 63
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 KKSGRNKALKKPEYLTM---EVLKKNLSLTPVEKNLHFIWGGQI----- 107
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 KLTINN--VDDPKFENLLVEAFKINGFTTNKNFIYQVLAIIHLGNITFTSMKSEQANF 382
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 -NDTAINYINQKQVNSDY-----TVKFVYDSNA----- 135
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 383 TNDSPIDRICELLAVDKOLFVQNLRLPKVKAGREFITKSKPNEKVFATDAFAYLYEKL 442
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 136 --FLINTLKKTIVES-----ATNLTLESFRENLD---PEFDYKPYRKRMEIIVDK 182
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 443 FQFIITKINEKLDQSHNGDDNTANNNDNNFFIGVLDIAGFEIFDINSPEQLCINVTNBK 502
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 QRHFIDYKYSQIEENPEFIIDNIITYLSNEYSKOLE-----ALNKYIEES-- 228
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 503 LQQFNNHH-SFILEQSEYLRNINNEFI--DFGQDLQPTIDLIETKQPMGILKLDEECL 559
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 229 -----LNKITANNNDIRNL--EKPADELDVRLYNQEL---VERWNLAASDILRI 274
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 560 MPKSDASPMKSLKNFTTHKFKSENKPKNGFIHHYAGKVEYNVENW-LQKNTDPISE 618
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 SMLKEDGGVYLDVILPGIQLPKS-----INKPDSITWTSWEMIKLEAIMKYKEYI 327
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 619 SIL-----NLLPDSQNELIADMFINDPHINRPO--TNGGNSKLKTASQKHQDL 666
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 328 PGYTKNFDMLDEEVQSPESALSCKSEIFPLDLDIKVSPLEVYKIAFANNVINQAL 387
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 667 KTLMDQ--LESTEPHFVRCILPNLEKANKFDKNLVGLQRC-----NGVLEGIR 714
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 ISLKDSYCSDLVINQIKRYKILNDNLNPSINEGTFDNTTKIFSDKLASINEONMMFM 447
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 715 IT-RAGYPNRMFDEFIQRYSIICNELSSPONKTNCEIIL-----KFVKLNPEDFKVGL 768
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 448 IKITVYLVGPADYRSTINLSGPGVYTCAYQDILLMFKDNSTNIHILLEPELRFPPKTK 507
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 769 TKI--FFKNGILGKLEIIRDLAKNIFTDLQK---VIRGNLRLVL-----KOK 812
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 508 ISQTEQEITSLWSFNQARAKSQFBEYKKGYPEGALGED-----DNLDFAQN---TVLD 558
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 813 IKEIQSAQIISRTWYTLDEIKNSPMWRLFFHVKPLEDSAKVLSKQKLENLQTLTVKL 872
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 559 KDYVSKKILSSMKTRNKRYIHIVQLQGDKISYEASCNLFSDKDPYSSILYQKNIEGSETA 618
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 873 KD--SEKLTGKLETDNEKLRQMNQLDEMINITSIAK--EKD---DKLSQLRLTENKSK 925
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 619 YYYVADAEIKEDKY-----RIPYQISNKRNIKLTFFIGHKSEFNT-----DT 662
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 926 HRIETLEIKLDPKQKQNDLINEHEKLTTRRSLELHDHKNKV-----EELNSLNKLHD 978
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 663 FANLDVDSLSSBIETILNLAKA-----DISPKYIEINLLGCNM----- 700
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 979 SAQSELSNLKQQLKEITKLNQAHSKELLKFKQWHDKSIEDLNEKVSRLNKLKQADLN 1038
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 701 FYSIYAETYPCKL---LLKIKDRVSELMPISIQDSITVSANQVVRINEEGKREILDH 757
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1039 SSKDVVISSE--HGKLQNEIMKWLKELCSDYERKQKDSQRTIDSLQIQKKEDYK----- 1090
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
QY 758 SGKWKNEESIKDSSKEYISFNPKNKIIVKSKYLHELSTLLQEI-RNNANSSDIDLE 816
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1091 -----SRVESKIEEAREKVTLKKGVEKKSQEIQYKSEIKKLAELASSNTKITQLDOS 1145
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 817 KKVMLTCEIINVASNDIROIVEGRIEERAKNLTSDSINYIKNEFKLIES-----ISDSLVDL 872
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1146 TKEL-----SSLKSNESKYVAE--IESAKQQLSKKI-----BEYSSIEADYNRLQTELKQM 1194
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 873 KHQGLDDSHFISPEDISKTEGPFRI-----RFINKETGNSIFETEKE-----IFSEVA 922
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1195 KQTN-----TEYSTRITELSTKLNQVTEAKSKIEKENQNPQPMFBEF 1239
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 923 TH-----ISKEISNIKDTIFDN-----VNGKLV-----KKVNL-----DAAHE 955
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1240 THMKLVNQEQTASLRKEKPEKPKLSELOMLKERIMNGSLTSMDLTPKRSIAIGDKSMI 1299
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 956 VNTLSAFFIOSLIEBYNTTKESL-----SNLSVAMKVQVYAOQLFSTGLNTI 1001
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1300 TNTVDS-----FNKEIENLKFQLOEQGNFORAENYATIELQKLNKLLTTRGLNTN 1350
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1002 TDASK 1006
DB 1351 TDYEK 1355
```

Search completed: November 5, 2003, 19:29:28
Job time : 36.75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:27:53 ; Search time 14.25 Seconds
(without alignments)
3028.567 Million cell updates/sec

Title: US-09-126-816B-6_COPY_1_1020

Perfect score: 5192
Sequence: 1 MNLVNAQLQKVMYVKFRIQ.....ITDASKVVELSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/prodata/1/iaa/5A COMB.pcp.*
- 2: /cgn2_6/prodata/1/iaa/5B COMB.pcp.*
- 3: /cgn2_6/prodata/1/iaa/6A COMB.pcp.*
- 4: /cgn2_6/prodata/1/iaa/6B COMB.pcp.*
- 5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pcp.*
- 6: /cgn2_6/prodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4086	78.7	2366	1	US-08-480-604A-10
2	4086	78.7	2366	2	US-08-405-496A-10
3	4086	78.7	2366	3	US-08-915-136-10
4	4086	78.7	2366	4	US-08-957-310-10
5	4086	78.7	2366	4	US-10-011-366-10
6	2783	53.6	2710	1	US-08-480-604A-6
7	2783	53.6	2710	2	US-08-405-496A-6
8	2783	53.6	2710	3	US-08-915-136-6
9	2783	53.6	2710	4	US-08-957-310-6
10	2783	53.6	2710	4	US-10-011-366-6
11	411.5	7.9	3169	4	US-09-453-702B-257
12	261.5	5.0	2184	4	US-09-417-485B-6
13	247.5	4.8	10182	4	US-09-134-001C-3159
14	237.5	4.6	3433	4	US-09-091-501B-10
15	235	4.5	3248	1	US-08-353-700-1
16	235	4.5	3248	5	PCT-US95-16216-1
17	224.5	4.3	1151	4	US-09-134-001C-3242
18	220	4.2	3696	4	US-09-134-001C-5080
19	218	4.2	2285	4	US-09-308-375-2
20	215	4.1	2662	4	US-09-595-684B-31
21	211	4.1	956	4	US-09-134-001C-4452
22	209.5	4.0	2504	4	US-09-328-352-5821
23	209	4.0	1053	3	US-09-315-793-52
24	208.5	4.0	1786	3	US-08-973-462-8
25	206.5	4.0	1010	4	US-09-134-001C-5178
26	204	3.9	3878	4	US-09-914-259-11
27	202.5	3.9	1169	4	US-09-255-829-20

28	202	3.9	840	3	US-08-974-549A-190	Sequence 190, Appl
29	202	3.9	872	3	US-08-851-843A-8	Sequence 8, Appli
30	202	3.9	872	3	US-08-851-843A-54	Sequence 54, Appl
31	202	3.9	872	3	US-08-974-549A-221	Sequence 221, App
32	202	3.9	872	3	US-08-854-050-8	Sequence 8, Appli
33	202	3.9	872	3	US-08-854-050-54	Sequence 54, Appl
34	202	3.9	872	4	US-09-430-323-8	Sequence 8, Appli
35	202	3.9	872	4	US-09-430-323-54	Sequence 54, Appl
36	200.5	3.9	859	1	US-08-053-614-2	Sequence 2, Appli
37	200.5	3.9	859	1	US-08-316-397B-2	Sequence 2, Appli
38	200.5	3.9	859	2	US-09-034-306-2	Sequence 2, Appli
39	200.5	3.9	859	3	US-09-259-437-2	Sequence 2, Appli
40	200.5	3.9	859	5	PCT-US93-09782-2	Sequence 2, Appli
41	200.5	3.9	1181	1	US-08-053-614-4	Sequence 4, Appli
42	200.5	3.9	1181	1	US-08-316-397B-4	Sequence 4, Appli
43	200.5	3.9	1181	2	US-09-034-306-4	Sequence 4, Appli
44	200.5	3.9	1181	3	US-09-259-437-4	Sequence 4, Appli
45	200.5	3.9	1181	5	PCT-US93-09782-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-10
; Sequence 10, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01763
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-10

Query Match 78.7%; Score 4086; DB 1; Length 2366;
Best Local Similarity 77.5%; Pred. No. 5.1e-237;
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;

QY 1 MNLVNAQLOKQVYKFRQEDYVAILNALEYHNHNSSESVVVEKYLKLDINNLTDNYL 60
DB 1 MSLVNRKQLEKMANVRFTQEDYVAILDALEYHNHNSENTVVEKYLKLDINSLTDIYI 60

QY 61 NTYKSGRNKALKKPKFYEYLTMEVLELKNLSLTVEKNLHFIWIGGQINDTAINYNQWD 120
DB 61 DTYKSGRNKALKKPKFYEYLTMEVLELKNLSLTVEKNLHFIWIGGQINDTAINYNQWD 120

QY 121 VNSDYTVKFDYDGNALINTLTKTIVESATNTLESFRENLDNPEFDYKFKRMEIY 180
DB 121 VNSDYNVYDGNALINTLTKTIVESAINDTLESFRENLDNPEFDYKFKRMEIY 180

QY 181 DKQHFIDYKSGIENPEFIIDNIIKTYSNEYSKDLALNKYIEESLNKITTANNNDI 240
DB 181 DKQKNFYNYKQAQREENPELIIDIVKTVLSNEYSKEIDELNTYIEESLNKIQNSGNDV 240

QY 241 RNLKFADEDLVRLNQVELVERNLAAASDILRISMLKEDGGVYLDVDPGLQDPLFKS 300
DB 241 RNFEFKNGESFNLYEQELVERNLAAASDILRISALKEIGGYLDVDMPLGQDPLFES 300

QY 301 INKPDSITNTSWEMIKLEAMTKYKVI PGVTSKNPDMDEEVORSFESALSSKSKSEIF 360
DB 301 IEKPSVTVDFWMTKLEAMTKYKVIPEYTSFHPDMDEEVORSFESVLSKSKSEIF 360

QY 361 LPDDIKVSPLEKIAFANNVINOALISLSDSYCSDLVINQIKNRYKILNNDNPSINE 420
DB 361 SSLGDMSEAPLEKIAFNSKGIINOGLISVKDSYCSNLIVKQIENRYKILNNSNPAISE 420

QY 421 GTDPNTMTKIFSKLASISNEDNMFMKIKNVLYKGFAPDVSTINLSGPGYITCAYOD 480
DB 421 DNDFTNTTTFDISIAENADNRFMELGKLVGFPPDVKTITNLSGPEAYAAAYQD 480

QY 481 LLMFKDNSTNIHLEPELNRFPKTKISOLTEQETISLWSFNQARAKQFEEYKGYFE 540
DB 481 LLMFKGSGNHIHLEADLNFEISKTNISQSTEQEMASLWSFDDARAKQFEEYKRYFE 540

QY 541 GALGEDNDLFAQNTVLDKDYVSKKILSSMKTRNKYIHYIVLQGDKISYEASCNLFSK 600
DB 541 GSLGEDNDLDFSQNVVDKEYLEKISLARSSRGYIHYIVLQGDKISYEAAACNLFAK 600

QY 601 DPVSSILYQKNIEGSETAYYYYADAEIKEDKIRPYQISKNKNTKLTFFIGHGKSEPT 660
DB 601 TPVDSVLQKNIEDSIAIYNNPGDGEIQEIDKYPISIDSPKLTFFIGHGKDEFT 660

QY 661 DTPANLDVSLSEIETILNLAKADISPKYIENLLGCNNFYSIYAEETYPGKLLKIK 720
DB 661 DIPAGFDVSLSEIEAIDLAKEDISPKYIENLLGCNNFYSINVEETYPGKLLKIK 720

QY 721 DRVSELMPSISODSITVSANQYEVRIEENEGREILDHSGKWNKESIIKDISSEYISF 780
DB 721 DKISELMPSISODSIIVSANQYEVRIENEGREILDHSGKWNKESIIKDISSEYISF 780

QY 781 NPENKIIVKSKYLHELSTLLQIRNANSDDIDLEKKVWLTECEINVASNIDRQIVEGR 840
DB 781 NPENKIIVKSKNLPSTLLQIRNANSDDIDLEKKVWLTECEINVASNIDTQIVEER 840

QY 841 IEBAKNLTSDSINYIKNEFKLIESISDSLYDLKHQNGLDDSHFISPEDISKTEGPRIRF 900

DB 841 IEBAKNLTSDSINYIKNEFKLIESISDALCDLKQNELEDHSFISFEDSETDEGSIRF 900

QY 901 INKBTGNSIFITEKEIFSEYATHISKEISNTKOTIFDNVNGKLVKKNLDAAEVNTLN 960
DB 901 INKBTGESIFVETKIFSEYATHITEISKIKGTIFDNVNGKLVKKNLDTTHEVNTLN 960

QY 961 SAFFIQSLIEVNTTKESLSNLSVAMKVQVYAOQLFSTGLNTITDASKVVELVSTALDETD 1020
DB 961 AAFIFIQSLIEVNTTKESLSNLSVAMKVQVYAOQLFSTGLNTITDAAKVVVELVSTALDETD 1020

RESULT 2
US-08-405-496A-10
Sequence 10, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEUROTOXIN
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSES: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08405496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-10

Query Match 78.7%; Score 4086; DB 2; Length 2366;
Best Local Similarity 77.5%; Pred. No. 5.1e-237;
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;

QY 1 MNLVNAQLOKQVYKFRQEDYVAILNALEYHNHNSSESVVVEKYLKLDINNLTDNYL 60
DB 1 MSLVNRKQLEKMANVRFTQEDYVAILDALEYHNHNSENTVVEKYLKLDINSLTDIYI 60

QY 61 NTYKSGRNKALKKPKFYEYLTMEVLELKNLSLTVEKNLHFIWIGGQINDTAINYNQWD 120

Db 61 DTVKSGRNKALKKFKYEVLTVELEKNNLTPVEKNLHFVWIGGQINDTAINYNQWKD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTLTESFRENLDNPEFDYKFKRMEIY 180
Db 121 VNSDYNVNVFYDSNAFLINTLKTIVESAINDTLTESFRENLDNPEFDYKFKRMEIY 180
Qy 181 DRQKHFDIYKYSQIENPEFIIDNIKTYSNEYSKDLALNKYIEESLNKITTANNNDI 240
Db 181 DRQKFNFINYYKAQRBNPELLIDDIIVKTYLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240
Qy 241 RNLEKPADDELVRNLQELVERNLAAASDIIRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFPKNGESFNLYQEQLVERNLAAASDIIRISALKKEIGMYLDVMDLPGIQDPLFS 300
Qy 301 INKPOSITWTSWEMIKLEAIMKYKEYIPGVTSKNFMDLDEEVSORSFESALSKSKSEIPE 360
Db 301 IKPSSSVTVDFWEMTKLEAIMKYKEYIPETSHFMDLDEEVSQSSFESVLASKSKSEIP 360
Qy 361 LPDDDIKVSPLVKIAFANNVINOQALISLKOSYCSDLVINQIKNRYKILNDNLNPSINE 420
Db 361 SSGDMEASPLEVKIAFNSKGIINOGLISVKOSYCSNLIVKQIENRYKILNLSNPAISE 420
Qy 421 GTDFTNTMTKIFSDKLASISNEDNMFMWIKITNYLVKGFAPDVRSTINLSGPGVYTGAYQD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMBELGKYLVRGVFPDVKTINLSGPEAYAAAYQD 480
Qy 481 LLMPKDNSTNIHLEPELBNFPKTKISQLTQEBITSLWSFNQARAKSQFEEKYKGYE 540
Db 481 LLMPKESGNIHLIEADLRNFESKTNISQSTEQEMASLWSFDDARAKAQFEEKYKGYE 540
Qy 541 GALGEDNDLDFACNTVLDKDYYSKTLSSMKTNRKEYIHYIYVLOQGDKISYASCNLPSK 600
Db 541 GSLGEDNDLDFSQNIYVDKEYLLEKISSLARSGERYIHYIYVLOQDKISYEAACNLPAK 600
Qy 601 DPYSSLYQKNIEGSTAYYYVYVADAEIKEDKIRIPYQIKNRNIKLTFIGHGKSEFNT 660
Db 601 TPYDSVLFOKNIEDSBIAYYVNPGEIYQIYKIPSIISDRPKIKLTFIGHGKDEFNT 660
Qy 661 DTFANLDVLSLSEIETILNLAADISPKYIEINLLGCNMFYSIYABETYPQKLLKIK 720
Db 661 DIPAGFDVPSLSEIEAADIADLAKEDISPKSIEINLLGCNMFYSINVEETYPQKLLKVK 720
Qy 721 DRVSELMPSISODSITVSANQYEVNRINSEGRILDHSGKWNKBSIISKSSKEYISF 780
Db 721 DKISELMPSISQDSIIVSANQYEVNRINSEGRRELLDHSGEWNKBSIISKSSKEYISF 780
Qy 781 NPKNKIIVKSKYLHSLTLQBIERNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840
Db 781 NPKNKITVKSKNLPESLTLQBIERNANSSDIIELEKVMLTECEINVISNIDTQIVEER 840
Qy 841 IBEAKULTSDSINYIKNEPKLIESISDLYDLKHQGLDSDHPISEFIDISKTENGPRIRP 900
Db 841 IBEAKULTSDSINYIKNEPKLIESISDLYDLKHQGLDSDHPISEFIDISKTENGPRIRP 900
Qy 901 INKGTNSFIETEKIFSEYATHISKEISNIKDTIFDNVNGKLVKVNDAHAHVNTLN 960
Db 901 INKGTESIFVETEKIFSEYANHIIEBISKINGTIFDVTNGKLVKVNLDTHVNTLN 960
Qy 961 SAFFIOQLSEYNTTKESLSNLSVAMQVQVYQALFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 AAFPIQSLIEYNSKESLSNLSVAMQVQVYQALFSTGLNTITDAAKVVELVSTALDETID 1020

RESULT 3

US-08-915-136-10
; Sequence 10, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.

Query Match 78.7%; Score 4086; DB 3; Length 2366;
Best Local Similarity 77.5%; Pred. No. 5.1e-237;
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;

Qy 1 MNLVKAQLOKQVYKFRIOEDYVAIILNALLEYHNMSESSVVEKYLKLDINLTDNVL 60
Db 1 MSLVNRKQLEKMANVFRFTQEDYVAIILDALEEHNMSENVTVEKYLKLDINSLTDIYI 60
Qy 61 NTVKSGRNKALKKFKYEVLTVELEKNNLTPVEKNLHFVWIGGQINDTAINYNQWKD 120
Db 61 DTVKSGRNKALKKFKYEVLTVELEKNNLTPVEKNLHFVWIGGQINDTAINYNQWKD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTLTESFRENLDNPEFDYKFKRMEIY 180
Db 121 VNSDYNVNVFYDSNAFLINTLKTIVESAINDTLTESFRENLDNPEFDYKFKRMEIY 180
Qy 181 DRQKHFDIYKYSQIENPEFIIDNIKTYSNEYSKDLALNKYIEESLNKITTANNNDI 240
Db 181 DRQKFNFINYYKAQRBNPELLIDDIIVKTYLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240

APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-10

```
QY 241 RNLEKFADEDLVRLNQEELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNLEKFADEDLVRLNQEELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSSKSKSEIF 360
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSSKSKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLKDSCDVLINQIKNRYKILNDNLNPSINE 420
Db 361 SSGDMEASPLEVKIAFANNVINOALISLKDSCDVLINQIKNRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKIFSDKLASISNEDNMFMKITYLVKGFAPDVRSTINLSPGVYTCAYOD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRMELGKYLVRGFFPDVKTINLSPGEAYAYQD 480
QY 481 LLMFKDGNSTNIHLPELNPFPKTKISQLTQEITSLWSFNQARAKSQFEYKGYE 540
Db 481 LLMFKEGSMNIHLIEADLNFEISKTNISQSTEQEASLWSPDARAKAQFEYKRYE 540
QY 541 GALGEDNLDFAQNTVLDKDYVSKKILSSMKTRKEYIHYIVOLQDKISYASCNLFSK 600
Db 541 GSYGEDNLDQSNIVVDKEYLEKISSLASRSEGYIHYIVOLQDKISYEAACNLFAK 600
QY 601 DPYSSILYQNGIEGSTAYVYVADAEIKEDIKYRIPYOISNKRNIKLTFIGHKSEFT 660
Db 601 TPVDSVLFQKNIEDSEIAYVYVADAEIKEDIKYRIPYOISNKRNIKLTFIGHKSEFT 660
QY 661 DTFANLDVDSLSSEIETILNLAADISPKYIEINLLGCMFYSIYASSETYPKLLKIK 720
Db 661 DIPAGFDVDSLSSEIETILNLAADISPKYIEINLLGCMFYSIYASSETYPKLLKIK 720
QY 721 DRYSELMPSTSDQSIIVSANQYEVNRNECKRILDHSGKWKINKEISIKDISKEYISF 780
Db 721 DKISELMPSTSDQSIIVSANQYEVNRNECKRILDHSGKWKINKEISIKDISKEYISF 780
QY 781 NPKENKIIVKSKYLHELSTLQIRNANSDDILEKKVWLTCENINVASNIDRQIVEGR 840
Db 781 NPKENKIIVKSKYLHELSTLQIRNANSDDILEKKVWLTCENINVASNIDRQIVEGR 840
QY 841 IEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHSPFISFEDISKTENGFRIRF 900
Db 841 IEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHSPFISFEDISKTENGFRIRF 900
QY 901 INKETSPIETETEKIFSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAAEVNTLN 960
Db 901 INKETSPIETETEKIFSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAAEVNTLN 960
QY 961 SAFFIOSLIBYNTTKESLSNLVAMQVYVQVYVQVYVQVYVQVYVQVYVQVYVQVYV 1020
Db 961 AAFIOSLIBYNTTKESLSNLVAMQVYVQVYVQVYVQVYVQVYVQVYVQVYVQVYV 1020
```

RESULT 4

```
US-08-957-310-10
; Sequence 10, Application US/08957310
; Patent No. 636518
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,310
FILING DATE: 23-OCT-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 24-OCT-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCES/DOCKET NUMBER: OPHD-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-310-10

Query Match 78.7%; Score 4086; DB 4; Length 2366;
Best Local Similarity 77.5%; Pred. No. 5.1e-237;
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;

QY 1 MNLVNAQLOKVVYKFERIQEDEVYVAIINALSEYHNMSSESVVEKYLKLDINLNTDNYL 60
Db 1 MSLVNRKQLEKVNVRFTQDEYVAILDALESEYHNMSSENTVVEKYLKLDINLNTDNYL 60
QY 61 NTKKSGRKNALKKFKKEYLTMEVLELKNNSLPVPEKNLHFIIWGGQINDTAINYNQWKD 120
Db 61 DTKKSGRKNALKKFKKEYLVTEVLELKNNSLPVPEKNLHFIIWGGQINDTAINYNQWKD 120
QY 121 VNSDITVKEVYDSNAFLINTLKTIVESATNTNLTESFRENLDNDPFDYKFKRMEIYY 180
Db 121 VNSDYNVNVFYDSNAFLINTLKTIVESATNTNLTESFRENLDNDPFDYKFKRMEIYY 180
QY 181 DKQKHFDYKSOIBENPEFIIDNLIKTYLSNEYSKOLEALNKYTEESLNKITTANNNDI 240
Db 181 DKQKFNIFYKQARENPELIIDNLIKTYLSNEYSKOLEALNKYTEESLNKITTANNNDI 240
QY 241 RNLEKFADEDLVRLNQEELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNLEKFADEDLVRLNQEELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSSKSKSEIF 360
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSSKSKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLKDSCDVLINQIKNRYKILNDNLNPSINE 420
Db 361 SSGDMEASPLEVKIAFANNVINOALISLKDSCDVLINQIKNRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKIFSDKLASISNEDNMFMKITYLVKGFAPDVRSTINLSPGVYTCAYOD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRMELGKYLVRGFFPDVKTINLSPGEAYAYQD 480
QY 481 LLMFKDGNSTNIHLPELNPFPKTKISQLTQEITSLWSFNQARAKSQFEYKGYE 540
Db 481 LLMFKEGSMNIHLIEADLNFEISKTNISQSTEQEASLWSPDARAKAQFEYKRYE 540
```


Db 958 VNTLNAFFIQSLIDYSSNKKVDLNDLSTSVKQVLAQFSTGLNTIYDSIQLVNLISNAV 1017

QY 1016 DETID 1020

Db 1018 NDTIN 1022

RESULT 7

US-08-405-496A-6

Sequence 6, Application US/08405496A

Patent No. 5919665

GENERAL INFORMATION:

APPLICANT: WILLIAMS, JAMES A.

TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

TITLE OF INVENTION: NEUROTOXIN

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/405,496A

FILING DATE: 16-MAR-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPND-01308

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-405-496A-6

Query Match 53.6%; Score 2783; DB 2; Length 2710;

Best Local Similarity 52.1%; Pred. No. 1.1e-158;

Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;

QY 1 MNLVNAQLOKQVYKFRQEBYVAILNALBENHNMSESSVVEKYKLDKNDLNTDYL 60

Db 1 MSLISKEELIKLAY-SIRPRENEYKTLTNDLDEYNKLTNNENKYLQKLKLNESIDVFM 59

QY 61 NTYKSGRNKALKKPKLEYLTMEVLEKNNLTPVEKNLHPITWGGQINDTAINYNOWKD 120

Db 60 NKYTSRRNALSNLKKOILKEVILIKNSNTSPVEKNLHPVWGGVSDIALEYIKQWAD 119

QY 121 VNSDYTVKPYVDSNAPLINTLTKKTIVESATNTNLTESFRENLDNPEFDYNKPKRMEITY 180

Db 120 INAEYNIKLWYDSEAFVNTLKAIVESSTTEALQLEEEIQNPQFDNNKPKYKCRMEITY 179

QY 181 DKQKHFIDYKSOIRENPEFIIDNIITKYLSEYKDLALNKYITEESLNKNTANNNDI 240

Db 180 DRQKFPINYKSOINKPTVPTIDDIKSHLVSEYNRDETIVLESYRTNSURKINSNHGIDI 239

QY 241 RNLEKFADEDLVRLYNQELVERWNLAASDIIRISMLKEDGGVYLDVILPGIQDLPFKS 300

Db 240 RANSLFTEQELLNIYSQELLNRGNLAASDI VRLALKNFGVYLDVDMPLGHSDFLFT 299

QY 301 INKPSITNTSWEMIKLEIAIMKYKEYIPOYTSKNFMDLDEEVORFESALSSEKSKSEIF 360

Db 300 ISRPSSIGLDRWEMIKLEIAIMKYKYNNTYENFDKLDQQLDKLQKLEIIESKSEKSEIF 359

QY 361 LPLDDIKVSPLEVKIAFAFANNVINQALISLSDSYCSDLVINQIKRYKILNDNLNPSINE 420

Db 360 SKLENLVSDLEIKIAFALGVSINQALISKQSYLTNLVIEQVKRYQFNLQHLNPAIES 419

QY 421 GTDFNTWKIFSDKLASISNEDNMFMKITNYLVKGPAPDVRSTINLSGPGVYTGAYOD 480

Db 420 DNNFTDTTKI PHDSLFGSATAENMFLTKIAPYLQVGFMPPEARSTISLSGPGAYASYD 479

QY 481 LLMFKDNSTNIHLLEPELNPEFPKTKISQLTEQISLWSFNQARAKSQPEYKKGYPF 540

Db 480 FINQENTIEKTLKASDLIEFKFPENNLSQLTEQISLWSFQDQASAKYQFEKYVRDYTG 539

QY 541 GALGEDDNLDPFAQNTVLDKDY-VSKKILSS--MKTRNKEYIHVIYLOQDKISYEASCNL 597

Db 540 GSLSEBDNGVDFNKNTALDKNLLNNKIPSNVVEAGSKNYHYIIQLQGGDLSYEATCNL 599

QY 598 FSKDPYSSILYQKIEGSETAYYYVAD--ABIKEIDKYRIPIQYISNKNINIKLTFIGHCK 655

Db 600 FSKNPKNSIIQORNM--NESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHCK 657

QY 656 SEFNTDTTANLVDLSLSEIETILNLAADISPKYIEINLLGCMFSPSYIYAETYPGKL 715

Db 658 DEFNTSEFARLSVDSLSENEISSFLDTIKLIDISPKNVEVNLGCMFSPSYFNVEETYPGKL 717

QY 716 LLKIKDRVSELMPSISQDSITVSANOYEVRINEEGKREILDHSGKWINKKEESTIKDISK 775

Db 718 LLSIMDKITSLDVPNKNISITIGANOYEVRINSEGRKELLASHGKWINKKEEAIMSLSK 777

QY 776 EYISPNKRNKIIVKSKYLHELSTLLQETIRNANSDDIDLEKKVMLTECEINVASNIDRQ 835

Db 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLDDASVSDTKFILNLLKLAIESSIGDY 837

QY 836 IVEGRIBERAKNLTSDSINVYKNEPKLIBSISLSYDLKHQGLDDSHFISPDISKTENG 895

Db 838 IYYEKLEPVKNIHNSIDDLIDEPNLENVSDLEYELKLNLDKYLISFEDISKNST 897

QY 896 FRIRINKETGNSIFETEKEIFSEYATHISKEISNIKDTIPDNVNGKLVKKNVLDAAHE 955

Db 898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGNLDNIQLDHTSQ 957

QY 956 VNTLNAFFIQSLIDYSSNKKVDLNDLSTSVKQVLAQFSTGLNTIYDSIQLVNLISNAV 1015

Db 958 VNTLNAFFIQSLIDYSSNKKVDLNDLSTSVKQVLAQFSTGLNTIYDSIQLVNLISNAV 1017

QY 1016 DETID 1020

Db 1018 NDTIN 1022

RESULT 8

US-08-915-136-6

Sequence 6, Application US/08915136

Patent No. 6290960

GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.

APPLICANT: THALLEY, BRUCE S.

APPLICANT: PADHYE, NISHA V.

APPLICANT: FIRCA, JOSEPH R.

APPLICANT: STAFFORD, DOUGLAS C.

TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/915,136

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/480,604

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496

FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLJA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPDH-01763

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-915-136-6

Query Match 53.6%; Score 2783; DB 3; Length 2710;

Best Local Similarity 52.1%; Pred. No. 1.le-158;

Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;

QY 1 MNLVNAQLOKMYVVFRIQEDYVAILNNALEYHNMSSESVVVKYKLDKNNLDNYL 60
DB 1 MSLISKEELIKLAY-SIRPRENEYKLTITLNDYKLTITNNENKYLQKKNESIDVFM 59
QY 61 NTYKSGRNKALKFKREYLTMEVLELKNLSLTPVEKNLHFIMIGGOINDTAIYNQWKO 120
DB 60 NKYTSRRNALNKKDKILKEVILKNSTSPVEKNLHFVWIGGEVSDIALEYIKQWAD 119
QY 121 VNSDYTVKPYVDNAFLINTLKKTVESATNNLTGPRENLNDPEFDYNNFKYKREMEIY 180
DB 120 INAEYNIKLYDSEAFVNLTKKAIVESSTTEALQLLEEIQNPQFDNMKFKYKREMEFI 179
QY 181 DKQGHFDYKYSQIEENPEFIIDNIITKYLNSNEVSKDLEALNKYIEESLNKITANNNDI 240
DB 180 DROGRFINYKQINQKPTVPTIDITKSHLVSENRDVTLESYRTNSLRKINSNHGDI 239
QY 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVLDILPGIQPLFKS 300

DB 240 RANSLFTEQELNLIYSQELLNRGNLAAASDIVRLALLKNFGVGLVDVMDPLGIHSDLPKT 299
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVORFESALSXSOKSEIF 360
DB 300 ISRPSSIGLDRWEMIKLEAIMKYKEYINNYTSNFENFDKLDQDKJQKFLIIESKSEIF 359
QY 361 LPLDDIKVSPLEVKIAFANNVINQALISLSDSYSDLVINOIKNRYKILNDNLPSINE 420
DB 360 SKLENLVSDLEIKIAFALGVSINOALISKQSYLTNLVIEQVKRYQFLNQHLPATBS 419
QY 421 GTDFNTMKIFSDKLASISNEDNMFMKITHYLVKGPAPVRSSTINISGPGVYTGAYOD 480
DB 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQGVEMPEARSTISLSPGAYASAYD 479
QY 481 LLMPKDNSTNIHLLEPELNPFPKTKISOLTEQBITSLWSFNQARAKSOFEEYKGYE 540
DB 480 FINQENTIEKTLKASDLTIEFKFPENNLSQLTEQBSINLSWSPDQASAKYQFEKYVRYDYG 539
QY 541 GALGEDDNLDFQNTVLDKDY-VSKKILSS--MKTRNKYIHYIYVLOQGDKISYEASCNL 597
DB 540 GSLSEDNGVDFNKNTALDKNYLLNKKIPSNVVEAGSKNYVHYIQLQGGDISYEATCNL 599
QY 598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKIDKYRIPYQISNKNIKLTFIGHOK 655
DB 600 FSKNPKNSIIQORNM--NESAKSYFLSDDDGESILELNKYRIPERLKNKEKVKVTFIGHOK 657
QY 656 SEFNTDTFANLVDVSLSSIEITILNAKADISPKYIEINLLGCNMFYSIYAEETYPGKL 715
DB 658 DEFNTSEFARLSVDLSLSNEISSFLDTIKLDISPKNVEVNLGCNMFYDFNVEETYPGKL 717
QY 716 LLKIKDRVSELMPSISODSITVSANQYEVINEEGKREILDHSGKWINKESIIKDISK 775
DB 718 LLSIMDKITSTLPDVNKSITIGANQYEVIRINSEKRELLAHSKWKINKERAIMSDLSK 777
QY 776 EYISFPKNENKIIVKSKYLHELSTLLQEIIRNNANSDDIDLEKKVMLTECEINVASNDRQ 835
DB 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLDDASVSPDTKFLNLLKNISSIGDY 837
QY 836 IVEGRIEAKNLTSDSINIKNEFKLIESISLSDVLDKHQGLDSDHSFTSFEDISKTENG 895
DB 838 IYYEKLEPVKNTIHNISIDDLDEFNLLNENSVDELVELKLNLDLDEKYLISFEDISKNNST 897
QY 896 PRIRPINKETGNSIFETEKEIFFSEYATHISKEISNICKDTIFDVGKLVKVKVNLDAAE 955
DB 898 YSVRFINKNGESVYVETEKEIFFSKYSEHITKEISTIKNSIITDVGNNLLDNIQDHTSQ 957
QY 956 VNTLNSAFFIQSLIEYNTTKESLSNLVAMKYQVYAQLFSTGLNTITTDASKVELVSTAL 1015
DB 958 VNTLNAAPFIQSLIDYSSNKNVDLNDLSTSVKQVLAQLFSTGLNTIYDSIQLVNLISNAV 1017
QY 1016 DETID 1020
DB 1018 NDTIN 1022
RESULT 9
US-08-957-310-6
; Sequence 6, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America

[illegible]

```

/ FILING DATE: 02-DEC-1993
/ APPLICATION NUMBER: US 07/985,321
/ FILING DATE: 04-DEC-1992
/ APPLICATION NUMBER: US 07/429,791
/ FILING DATE: 31-OCT-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ingolia, Diane E.
/ REGISTRATION NUMBER: 40,027
/ REFERENCE/DOCKET NUMBER: OPND-01121
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 705-8410
/ TELEFAX: (415) 397-8338
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2710 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6

Query Match 53.6%; Score 2783; DB 4; Length 2710;
Best Local Similarity 52.1%; Pred. No. 1.1e-158;
Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;

QY 1 MNLVKAQLQKVVYKFRQEDYVAI LNALBEYHNMSSESVVEKY LKLDNNLNDYL 60
DB 1 MSLISKEELIKAY-SIRPRENEYKYLTDNDEYKLTNNENKYLQKKLNESIDVPM 59
QY 61 NTYKSGRNKALKKKPKXYLTMEYLELKNNSLTPVEKNLHPFIWIGQINDTAIYNQWKD 120
DB 60 NKYKTSRRNALSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGVSDIALEYIKQWAD 119
QY 121 VNSDYTVKVFYDSNAFLNLTWLTKTIVESATNTLESFRENLDNDPEPDYKFKRMEIYY 180
DB 120 INAEYNIKLYWDEAFVNLTKKAIYESSTTEALQLLEBEIQNPQPDNNKFKYKRMETFY 179
QY 181 DKQKHFDYVYKSGIENPEFIIDNIITKYLNSYKDLALNKYIEESLNKJITANNNDI 240
DB 180 DRQGRFINYKSGIQNPTVPTIDDIKSHLVSYNRDETVELSYRNSLRKINSNHGIDI 239
QY 241 RNLEKFAEDLVRLYNQELVERNLAAASDILRISMLKEDGGYVLDVILPGIQDLPKFS 300
DB 240 RANSLFTEQELLNIYSQELLNRCNLAAASDIVELLALKNFGVYLDVDMPLGHIHSLFKT 299
QY 301 INKPSDITNSWMIKLEAIMKYEYIPGYTSKNFMDLBEVORSEFALSSEKSKSEIF 360
DB 300 ISRPSSIGLDRWEMIKLEAIMKYKVINNYTSENFDKLDQKDNPKLIIESKSEKSEIF 359
QY 361 LPDDIKVAPLEVKIAFANNSVINOALISIKDSCYSDLVINOIKNRYKILNDNLPSINE 420
DB 360 SKLENLVSDLEIKIAPALGVSINOALISKQGSILTNLVIEQVKRYQFNLQHLNPAIES 419
QY 421 GTDFNTMTKIFSDKLASIGNEDNNMFMIKITNLYKVGFAFPDVRSTINLSGPGVYTGAYOD 480
DB 420 DNNFTDTTKIFHDSLFNSATAENSMLTKIAPLVQGVFPEARSTISLSPGAYASAYD 479
QY 481 LMFKNSTNIHLLPELRFNPEPKTKISOLTEQETSLWSFNQARAKSQPEYKKGYPE 540
DB 480 FINLQENTIEKTLKASDLIEFKPEPNLSQLETEINSLWSFPQASAKYQFEKYVVDYTG 539
QY 541 GALGEDNDLDFQONTVLDRXDY-VSKILSS--MKTRNKEYIHVIVLQGDKISYEASCNL 597
DB 540 GSLSENGVDNFNKTALDRNLYLNKIPSNVNEEAGSKNYVHYIIQLQGDIDISYEATCNL 599
QY 598 FSKDPYSSILYQKNEGSETAYYYVAD--AEIKEIDKYRIPYQISGNKRNKILFTFGHGK 655
DB 600 FSKNPKNSIIIQRM--NESAKSYFLSDGSEILELNKYRIPERLNKKEKVKVTFFIGHGK 657
QY 656 SEFNTDTFANLVDVSLSSSETETILNLAKADISPKYTEINLLGCMNMFSEIYAEETYPGKL 715
DB 658 DEFNTSEFARLSDVLSNIESSFLDTIKLDISPKVNEVNLGCMNMFSDYFNVEETYPGKL 717

```

```

QY 716 LLKIKDRVSELMPSISODSITVSANQYEVIRINEEGRKILDSHGKWINKEESIIKDISK 775
DB 718 LLSIMDKITSTLPDVNKNISITIGANQYEVIRINEEGRKELLAHSGKWINKEEAIMSDLSSK 777
QY 776 EYISNPKNKLIYKSKYLHELSTLLQOIRNANSSDIDLEKKVWLTECEINVASNIDRQ 835
DB 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLLDASVSPDTKFTLNNLKLNISSIGDY 837
QY 836 IVEGRIBEAKNLTSIDNINIKNEFKLIESISDSLYDLKHQGLDDSHFISPEDISKTEG 895
DB 838 IYYEKLEPVKNIIHNSIDDLIDEFNLLNENSVDELVELKLNLDDEKYLISFEDISKNST 897
QY 896 FRIRINKETGNSIFETIEKEIFSEYATHISKEISNIDKTIPTDNVNGKLVKKNVDAAE 955
DB 898 YSVRFINKNGESVYVETEKEIFSKYSEHITKISITIKNSIITDVNGLLDNILQDHTSQ 957
QY 956 VNTLSAFFIOSLIENYNTTKSLSNLSVAMKVQVYLAQLFSTGLNTITDASKVVELVSTAL 1015
DB 958 VNTLNAAFFIQSLIDYSSNKKVDNLSTSVKQVLAQLFSTGLNTIYDSIQLVNLISNAV 1017
QY 1016 DETID 1020
DB 1018 NDTIN 1022

RESULT 11
US-09-453-702B-257
; Sequence 257, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Berna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 257:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3169 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 257:
US-09-453-702B-257

```

Query Match 7.9%; Score 411.5; DB 4; Length 3169;
Best Local Similarity 20.9%; Pred. No. 3.6e-16;
Matches 257; Conservative 195; Mismatches 453; Indels 323; Gaps 54;

QY 14 YKPRI---QEDYVAILNAL---EYHNMSESVVEKYLKLDINNDNYLNTYTKSGR 68
DB 160 YIKIRKTRGAEDOTTITQSLIINELLNGVDRTTI--PPQKISELNDIITHSYENMQIKNSR 218
QY 69 N--KALKKPKFYLWTEVLELKNS-----LTPVEKNLHFTWIGQ 106
DB 219 KGIELVKGCELLSSLIINDKNGKQSDNASKIINLLGIEYQSHKVDIEPFIHAWVAGA 278
QY 107 INDTAINYNQKVDNSYTVKPYVDSNAF-----LINTLAKT----- 144
DB 279 PDNTFSYITAFNLTYKDYLLWIDPNAFGAAPSGILKNAMVAMRLRTRTPHLAE 338
QY 145 -----IVESATNTLE---SFRENLDPEFY-----NKFYRKMEIYYDKQKFI 187
DB 339 EMNEVLKIQNTQNETIEPKETRERLKELENRYKSLTSEKPKNFVFFLESMIGHQDNFY 398
QY 108 DYKSOIENPEFI-----IDNIK--TVLSNEYSKOLEALNKYIEESLNKITTANN-- 238
DB 399 TYCISNGISNTDISELDELTVNLKLSPEVQDFKSTVEKNKRDIDLKNTISQKFGDRF 458
QY 239 ---DIRNLEKFADEDLVRLYNQELVERNMLAAASDILRISMLKEDGGVYLDVILPGIQP 295
DB 459 QLRDINTLESFKQPDYPPYQEMLLRWYAAASQVRINILKEYGGIYDTDILPAYSD 518
QY 296 DLFPKINKPDSITNTSWEMIKLEAIMYK--BYIPG--YTSQVMDLBEVORSFESALS 351
DB 519 KVSQIINE--KSDKRFPEFLKRLRISELSILKGEKYSIKH--DGLDETTLNQLNNIL- 575
QY 352 SKSDSEIFPLDDIKVSPLEVKIAPANNVI-----NOALISLKDSVC 395
DB 576 SIEK-----LTDDY--FRVETKVVRDTPKI--FRVQKTWTNTWIRGNMFLMTHGSKC 630
QY 396 SLDVINQIKRYKIILNDNLNPSINEGTDFNTMKIFSDKLASISNE-----DNM 444
DB 631 IDPILSGQKKQYLEL--QRIRDNISYNLFPYTT-----EDLSLNNVAIGGIPAKKYLEHG 684
QY 445 MPMIKITNLYKVGPPADVRSTINLSGP----- 471
DB 685 LF-----SEYRQDGTPIYVYSTLISGPDIMRMQMKYKSLGRIGEVHIKONKLSDVNPL 740
QY 472 GYVTGAYQDLMF---KONSTNIHLEPE-----LRNFPFKTISOLT-----BOEITS 518
DB 741 GYASNKDKNSFNWLPVSGVINDITPDDESSWAVRNDINDIKILPEKINCHVPEKLPTS 800
QY 519 LMSFNQARAKSQFEYKKGYPGALGEDDNDLDFAQNTVLDKDYVSKKILSSMKTNRKBYI 578
DB 801 LY-----YEIDSRSPQGW---DNKSIKHVTEINKDLI--KOINLLLTSSNDV 844
QY 579 HYIVLQGGKISYEASCNLFKDPYSSILYQKNIEGSETAYYYYVADAEIKEDIKRIPY 638
DB 845 KLLIKL--DRELYAISSKI-----DNPLALSIRTLQLANLYVTNTFEPENTINFTY 896
QY 639 QISNKRN-----IKLTFIGHCKSEBNTDTFANLDVS-----LSSEIET 677
DB 897 DYPRKQDDLLSAIKL-----FSRNDADTKIIWVNSVMEKNVFLREVISC 942
QY 678 IINLAKADISPKYIEINLLGCMNFSYIABETYPKGLLLKIKDRVSELMPSISQDSI-- 735
DB 943 VLRSKKVD--SYINEN-----KKNLSKEDA--GALRDYAKLAKMKELPSMLDDGGYCK 990
QY 736 TVSANQYEVRIINEEGREILDHSGKWINKEESIIKDISKEYISFNP-----K 783
DB 991 IITTNAY-----IKERDKL-----SGIYNIENSIISGHESFDIIRSNQHEWGLSTVEQPK 1042
QY 784 ENKIIVKSYKHELSTLLOEIRNNANSSDIDLEKKVMLTECHINVAS-----NIDRQIVE 838
DB 1043 KPEFVYKSE--LSSAKSIFDDIKK--YITDPETKRNVLVHQLDSDIKERIAFLDISHYAVP 1100
QY 839 GRIEBAKNLTSINSYKNEFKLIESI-----SDSLYDLKHQNGL 878

DB 1101 GSLLEKLQLS---GYVPSDINIITAEYLLASGVSHGYVVPAPSDKLELLRRHTK 1156
QY 879 DDSHFISPEDISKTEGPRIRPINKETGNSI---PIETEKEIFPSYATHISKEISNIKDT 935
DB 1157 SNSEWI---EKITP---YVYDILSDNVSNLVRPPLSEQKKILNDIKLEISKVS---EQ 1207
QY 936 IFDYNVNGKLVKKNVLDAAHEVNTLNSAFFIOSLIEYNTTK-----ESLSNLSV----- 983
DB 1208 YPMKLTEQSSVIGIKYSVDFRYNENLFLSLPINQNLTLPMRYFEMLYDIHIGIEN 1267
QY 984 -AMKVQVYAQFSTGLNTITDASKVVEL 1010
DB 1268 KANREFIYSKFFSSLMLDLFLINDERVLNL 1295

RESULT 12
US-09-417-485D-6
; Sequence 6, Application US/09417485D
; Patent No. 6541202
; GENERAL INFORMATION:
; APPLICANT: Long, David M.
; APPLICANT: Metz, Anneke M.
; APPLICANT: Love, Ruschelle A.
; TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
; FILE REFERENCE: 47714-5009-US
; CURRENT APPLICATION NUMBER: US/09/417,485D
; CURRENT FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2184
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (330)..(335)
; OTHER INFORMATION: Xaa at position 330 = Leu or Ile;
; OTHER INFORMATION: Xaa at position 335 = Asp or Gly.
US-09-417-485D-6

Query Match 5.0%; Score 261.5; DB 4; Length 2184;
Best Local Similarity 19.7%; Pred. No. 2.2e-07;
Matches 252; Conservative 193; Mismatches 442; Indels 395; Gaps 61;

QY 10 QKVVYKFERIQEDEVAILNALEEV--HNMSES-----SVVEKYLKLDI-----NNLTD 57
DB 572 KKCIPILGSKHNFKIFLKNVYKESFSLNQVMKNIKVKNIFQKKISKYKN 631
QY 58 NVL--NTYKSGRNKAL--KPKFELYT---MEVLEKNKSLTPVEKNLHFTWIGQIND 109
DB 632 RILLKNIFDNNYENKILHRNNKEIITNDINKIYKNKNDNL-----NNSP-----KIKT 681
QY 110 TAIN-----YINQWKDVN-----SDYTKPVYDSNAPLINTLUKTIVESATNTL-- 154
DB 682 TLFNKLRKRYFNKIKKINIAIQKRLMNRILVFLF--NVFIMPLIRRFPLTKSBOTLHK 739
QY 155 -----ESFR-----ENLNDPEFY--NKP-YRKRMEIYYDKQKH 185
DB 740 TIFDRKIWNHFTKISNFCYHQIFRNKGLKGRNEPKMDYQNMFNVKKGSKI--KTNK 797
QY 186 FIDYVYSQIENPEFIIDNIKTLYLSNEYSKOLEAL---NKYIEESLNKITTANN---GN 238
DB 798 YIFIKMKKKSTKCNKNNKFSKCPKPKKKKNLYNITRHNNIFIKKDEKSKTNLNLK 857
QY 239 DIRNLEKPAD--EDLVRLYNQEL-----VERNMLAAASDILRISMLKEDGGVYLDVILPGI 293
DB 858 SIDNLYKLEINKKSVRPYIKKFFYKIKKKYFALKKWKVIHMRMAKEE----- 904
QY 294 QPDLEKSNKPDSTINTSWEMIKLEAIMKY-----KEYIPGYTS-----KN 334
DB 905 -----KS-----NIKLERAPGFHFIFAQEKHEILKYFSSHPFQNRKINYGR 946

QY 794 LHELSTLQIRNAN-----SSDIDLEKKVMLTECEINVASNIDROI VEGRIEBAKVL-T 848
DB 7459 LNSNMQLQIIVNDNTVQNSDF-----INEDSS-QQDAYWHAQAAKDLIT 7505
QY 849 SDSINYIKNEP-KLIESISDSLYDLKHQGLDSDHPISPIDISKTBNGPRIRPINKETG- 906
DB 7506 AHTPTMDKQIDQAIENIKOALNDLHGSNKLSB-----DKCEASBQLQNLNSLTNG 7556
QY 907 -----NSIF-LETEKEIESEVATHISKEISN-----IKDTIFDYNVNGKLVKKNLDAHE 955
DB 7557 QKDTILNHIPSAPTRSQVEKTIAS--AKOLNNTMKALRDSIADNNEILOSSKYFNDSBQ 7614
QY 956 VNTLSAF-----FIQSLIEYNTTKESLSNLSVAMKV---QVYAQLF 994
DB 7615 QYANQAVNKAANIINDQTPVMADEIQSVL--NEVQTKONLHGDKLANDKTDQAAT 7672
QY 995 STGLNITDASK 1006
DB 7673 LNALNYLNAQR 7684

RESULT 14

US-09-091-501B-10
; Sequence 10, Application US/09091501B
; Patent No. 6518413
; GENERAL INFORMATION:
; APPLICANT: Tinsley, Jonathon M
; APPLICANT: Davies, Kay E
; TITLE OF INVENTION: Utrophin gene expression
; FILE REFERENCE: 620-42
; CURRENT APPLICATION NUMBER: US/09/091,501B
; CURRENT FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: PCT/GB96/03156
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: GB 9525962.8
; PRIOR FILING DATE: 1995-12-19
; PRIOR APPLICATION NUMBER: GB 9615797.9
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: GB 9622174.2
; PRIOR FILING DATE: 1996-10-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 3433
; TYPE: PRT
; ORGANISM: Artificial Sequence
; NAME/KEY: misc feature
; LOCATION: (239) ... (250)
; OTHER INFORMATION: Description of Artificial Sequence: Full length
; OTHER INFORMATION: utrophin construct; Xaa = unknown
US-09-091-501B-10

Query Match 4.6%; Score 237.5; DB 4; Length 3433;
Best Local Similarity 18.8%; Pred. No. 1.1e-05;
Matches 227; Conservative 212; Mismatches 450; Indels 321; Gaps 57;
QY 27 ILNAAEYHNMSSESVVEKYLLKDINNLTNDYNTYKSGRNKALKKFKYLTMEVL-E 85
DB 699 LLNLWLKWTAIQTTEIKETKMKQDTSSEMKK-LKALEKEQREIRIPRADELNGTGQILVE 757
QY 86 LKNSLTPVEKNLHPTWIGGQINDTAINYNQWKNVS-----DYTVKVFYDSNAP---- 136
DB 758 QMGKEGLPTE-----EIKNVLEKVSSEKWNVSQHLIEDLERKIQLOEDINAYFKQL 807
QY 137 -----LINT-----LKKTIVESATNTLESPE-----NL----- 161
DB 808 DELEKVIKTEKWEVKHTSSESRSQSLDSKQCSQRELTLLGLHPKIEWARASCALMS 867
QY 162 --NDPEFDYNKF-----YKRMEEIYDKQHFIDYKSO-----IENPEFIIDNI 205
DB 868 QPSAPDFVQRGPFSLGRYQAVQEAVEDRQOHLLENELKGPQGHAYLETTLKTLKDVLNDS 927

QY 206 IKTYLSNEYSKDEALNKYIEB--SLNKITANN-----GNDIRNLEKFADEDLVRLYN 256
DB 928 NKAQVSLVNLNDLAKVEKALQEKTKTILDEILENQKPAHKLAEETKALEKVPDVEKLYK 987
QY 257 QSLVE---RWN---LAAASDI-----LRISMLKEDGCV---YLDVDLPGIQDPLFKS 300
DB 988 QEFDDVQGWKVKLVLSKDLHLLEBEIALTLRAFEADSTVIEKWM-----GVKDFLMQ 1042
QY 301 -----INKPDSITNTSWEMIKLEAIMK-----YKEYIP-----GYT 331
DB 1043 QAAQGDAGLQQLQDOCSAFVNEIETISSLKNKMEIETNLSRSGPVAGIKTWQTRLG 1102
QY 332 SKNFOMLDEEV-----QRSFESALSSKSKSEIFPLDDDIKVSPLYEVKIAFANNVIN 384
DB 1103 QTQLEKLSKEIATQKSRLSBSQKAAKALKDLAEQWMTQAEVEELDERFEVKSPEELE 1162
QY 385 QALISLKOSYCSDLVINQIKRYKILNDNLNPSINEGDTFTMTKIFSKLASISNEDNM 444
DB 1163 SAVEEMKRAK-EDVLQKEV--RVKILKONI-----KLLAAKVPSSGGQELTS 1205
QY 445 MFWIKITNYL-----KVGFPDVRSTINLSGPOVYTGAYQDILLMPKD-NSTNIHLE 495
DB 1206 ELNVVLENYQLLCNRIRGKCHTLEEVMS-----CWIELHLYDLLETTWLNTLE 1253
QY 496 PELRNFPKPKTISOLTEQEIITSLMSF--NOARAKSQPEEYKKGYPGEGALGEDDNDLPAQ 553
DB 1254 ERKSTEVLPKTDVNE-ALESLSVLHPADNRITQIRELQTLIDGG----- 1301
QY 554 NTVLDKDYVSKKILSSMTNRKEYIHYIVQLQDKISYBASCNLFPSKDPYSILYOKNIE 613
DB 1302 --ILD-DIISEK-LEAFNSRVEDLSHL---ASSKQISLEKQLQVLRRETQMLQVLQESLG 1354
QY 614 GSETAYYYVADAIEIKDKYRIPQIS-----NKNIKLPTFGCHK 655
DB 1355 ELDKQLTLYLTD---RIDAFQVQEAQKIQAIEISAHELTLEBLRNNMSQPLT----- 1404
QY 656 SEFNFTDTPANLDVDSLSSSEIETILNKAADISPKYIEINLLGCMNPSYSIYABETYPGKL 715
DB 1405 SPESRTARGSQMDVLQKLRVSTYKFLQFQKPAFQRMLODKRVLDGVKAE----- 1457
QY 716 LLKIKDRVSELMPSISQDSITVSANOYEVRIINEBGRKREILDHSGKWINKEESI-K--DI 772
DB 1458 -LHVLD-VKDDVPDVIQTHLDKMKLYKTL--SEVKLEV-----ETVIKTGRHI 1502
QY 773 SSKEYISFNPKENKIIVKSKYLHELSTLLOETRNANSSDIDLEKKVMLTECEINVASNI 832
DB 1503 VQKQQT-D-NPKG-----MDEQLTSLKVLNDILGAQVTEGQDLERASQLARKMKKEAASL 1556
QY 833 DRQIVBGRJEEAKNLTSD-----SINYIKNEFKLIE-----SISDSLYDLKHQNG 877
DB 1557 SEWLSATETELVQKSTSEGLDGLDTEISWAKNVLDLEKRAKADLNTITESSAAL--QNL 1614
QY 878 LDDSHFISPIDISKTBNGF-RIRFINKETGNSIF-IETEKIIFSEYATHISKEISNIKOT 935
DB 1615 IEGSEPILEERLCVLNAGWSRVRTWEDMCNTLMNHQNLQEIFDGNVAHISTWLQ-ABA 1673
QY 936 IFDNVNGK-----LVKKV--NLDAH--EVTNLNSAFFIOS-----LIEY 971
DB 1674 LLDEIEKPTSQEIEIVKELVSELDDANLQVENVRDQALILMNARGSSRELVEPRLAEL 1733
QY 972 NTKTESLS-----NLSVAM---KVQVYAQLFSTGL-----NTITDASKVVE--L 1010
DB 1734 NNNFEKVSQHIISAKLLIAQEPLOYCLVTTFETGVPFSDLEKLENDIENMLKFVEKHL 1793
QY 1011 VSTALDETID 1020
DB 1794 ESSDEDEKMD 1803

RESULT 15
US-08-353-700-1
; Sequence 1, Application US/08353700

Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J.
APPLICANT: RATNER, JEROME B.
TITLE OF INVENTION: NUCLEIC ACID ENCODING A
TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
US-08-353-700-1

Query Match 4.5%; Score 235; DB 1; Length 3248;
Best Local Similarity 20.4%; Pred. No. 1.5e-05;
Matches 254; Conservative 192; Mismatches 441; Indels 360; Gaps 63;

QY 1 MNLVNAQLOKQVYVPRIOEDYVAILNALEYHN-----MSSESVVEKYLK---- 48
DB 431 MNLVNAQLOKQVYVPRIOEDYVAILNALEYHN-----NNLEEFKQKLCRAEQAFQASQIKENELRRSME 482
QY 49 -LKDINNLTNLYNTYKSGRNKA-----LKKFKEYLTM-----EVLELKNNSLTP 93
DB 483 EMKKNL-----LKSHSEQKAREVCHLEAEKKNIKQCLNOSQNFAMKAKNTSQET 535
QY 94 VEKNLHFIWIGQIND-----TAINYNOMKDVNSDYTVKVFVYDSNAFLINTLK 142
DB 536 MLRDLQ-----EKINOENSLTEKULAVADLEKORDCSQDLLKREHHIEQ-LNDKLS 589
QY 143 KTIVES-ATNNTLESFRENLDPEFDYNYKRYKRMBEIIYDKQKHFDIYKS-----QI 194
DB 590 KTEKESKALLSALE-----LKKKEYBELK-----EETKLFSCWSENEKLLTQM 633
QY 195 EENPEFI-----IDNIKT--VLSNEYSKDLKALNKYIEESLNKINTANGNDIRNLEKF 246
DB 634 ESEKENLQSKINHETCLTKTQQIKSHSEYNERVTRLEMDRE-----NLSVEIRNLHN 685
QY 247 ADEDLVRLYNQEL--VERWNLAASDILRISMLKEDGVVYLDVILPGIQPDLFKSINKP 304
DB 686 LDSKSVSEVETOKLAYWELQKAEFSD---QKHQKEIENMCLKTSQITGGVEDLEHKL--- 739
QY 305 DSITNTSWEMIKLEATMK---YKEYIPGYTSKNFMDLDEBQVRSFESALSSKSDKSEIFL 361
DB 361 DSITNTSWEMIKLEATMK---YKEYIPGYTSKNFMDLDEBQVRSFESALSSKSDKSEIFL 361

Search completed: November 5, 2003, 19:33:20
Job time : 20.25 secs

Db 740 -----QLLSNEIMDKDRCYQDLHAEYESLR-DLL-----KSKDASLVTNEDHQRSLL 785
QY 362 PLDDIKVSPLEVKIAFANNVINOALISLKDSVCSLVINQIKRYKILN--DNLNPSIN 419
Db 786 AFDQ-----QP-AMHSPA-NIIEGQSPSESECRLEADQSPKNSAILQNRVDSLEFSLE 840
QY 420 EGTDFNTMKIFSDKLASIN--EDNMFMKITNYLVKGFADPDRVSTTNLSGPGVVTGA 477
Db 841 SQQMNSDLQKQCEELVQIKGIEENLM-----KAEQHQSFVAETSORISKLOED--TSA 894
QY 478 YQDLMMFKDONSNIHLLPELRFNFPKTKISQLTQETISLWSFNQARAKSQFEEYKKG 537
Db 895 HQNVV-----AETLSALENKEKELQLLNDKVE--TEQ-----ABIQELKKS 933
QY 538 --YFEGALGEDDNLDPQNTVLDKDVYSKILLSMKTNRKEYHYHYVQOGDKISYEASC 595
Db 934 NHLLEDSLKE---LQLLSETL-----SLEKEMSSIIISLNKRETEELTOENGTUKLEINASL 986
QY 596 NLFSDKPYSSILYQKNIEGSETAYVYVADAIEKEI-DKYR-----635
Db 987 NOEKMN-----LIQKS-----ESFANYIDREKSISELSDDQKQEKLLILORCEETGNAYED 1038
QY 636 --IPYQISNKRNIKLTFI-----GHGKSEFNTD--TFANLDDVDS 670
Db 1039 LSQYKAAQEKSKLECLLNECTSLCENRKNLEQLKEAFKEHQEFLLKLAFAERNQN 1098
QY 671 LSSEETIILNLAADISPK-----YIEIN-----LLG 697
Db 1099 LMLELTVQALRSEMTDNQNNKSSEAGGLKQBIIMTLKEEQNMQKEVNDLLQENQLMK 1158
QY 698 CNMFSYIYAETYPQKLLKIKDRVSE---LMPISQDSITTVSANQY-----742
Db 1159 VMKTRHECONLESEP--IRNSKRESERNQCNFRQDMOLEKVEISLDSYNAQLVQLEAM 1216
QY 743 ----EVRINE-EGRBILDPHSGKI--NKEESIIDISSKEYISFNPKNKIIVKSKYL- 794
Db 1217 LRNKELKQSEKEKECLOHELQITRGDLETSLNLDQMDSQEIISGL--KCEIDAEEKYIS 1274
QY 795 --HELST-----LLQETRNANSSDIDLEKKVMTETEINVASNID--RQIVEGR 840
Db 1275 GPHELSTSQNDNAHLQCSLQTTMNKLN-----ELEK-----ICEILOAEKYELVTELNDNR 1325
QY 841 IE-EAKNLTSDSINYIKNEFKLIESDSLV-----DLKHQNG---LD 879
Db 1326 SECITATRKMAEVEGKLLNEVKILNDDSGLLHGELVEDIPGGEFGQPNQHPVSLAPLD 1385
QY 880 DSHFISFEDISKTENGFRIRFINKETGNSIFITEKEIFSEYATHISKEISNIK-----933
Db 1386 ESN--SYEHLTSLDKVEQMHFAELQE-KFLSLQSEHKILHDQHCQMSKMSLQTVYVDSL 1442
QY 934 -----DTIFDNVNGKLVKYNLDA-----AHEVNTLNSAPFIQSLEY 971
Db 1443 KAENLVLSNLRNFOGDLVKENQGLGLEGLVPSLSSSCVPSDSSSLSGDGSFYRALLEQ 1502
QY 972 NTTKESLSNLSVAMQVQVYA--OLFSTGLNTITDASKVVELV-STAL 1015
Db 1503 TGDMSLLSNEGAVSANQCSDEVFCSSLQTVYVDSLKAENLVLSNL 1549

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:31:53 ; Search time 24.375 Seconds
(without alignments)
7187.047 Million cell updates/sec

Title: US-09-126-816B-6_COPY_1_1020
Perfect score: 5192
Sequence: 1 MNLVKNKAQLQKVMYVKFRIQ.....ITDASKVVELVSTALDETID 1020

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				Published Applications AA:*	
1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*	2:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*	3:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*	5:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*	6:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*	8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*	9:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*	11:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*	12:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
13:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*	14:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*	15:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*	17:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*	18:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4086	78.7	2366	15	US-10-011-366-10
2	2793	53.6	2710	15	US-10-011-366-6
3	411.5	7.9	3169	15	US-10-114-170-257
4	261.5	5.0	2184	12	US-10-304-095-6
5	256	4.9	1948	12	US-10-032-585-7611
6	249	4.8	1639	15	US-10-087-464-10
7	246	4.7	1881	12	US-10-032-585-7646
8	224	4.3	1270	12	US-10-032-585-7127
9	221	4.3	1805	11	US-09-820-843A-73
10	218	4.2	2285	10	US-09-932-183A-2
11	213	4.1	1979	15	US-10-205-823-419
12	212	4.1	1610	15	US-10-155-533-9
13	209.5	4.0	993	9	US-09-815-242-5809
14	209.5	4.0	1002	9	US-09-815-242-12899
15	209.5	4.0	1002	9	US-09-815-242-13158

16	209.5	4.0	1240	12	US-10-032-585-7366	Sequence 7366, Ap
17	208.5	4.0	1786	10	US-09-742-096-3	Sequence 3, Appli
18	207	4.0	996	9	US-09-815-242-5251	Sequence 5251, Ap
19	207	4.0	1009	9	US-09-815-242-12141	Sequence 12141, A
20	205.5	4.0	696	12	US-10-032-585-7808	Sequence 7808, Ap
21	204	3.9	1171	12	US-10-032-585-7519	Sequence 7519, Ap
22	204	3.9	3878	12	US-10-080-608A-11	Sequence 11, Appl
23	202.5	3.9	1169	12	US-10-241-596-20	Sequence 20, Appl
24	202	3.9	872	10	US-09-843-676-8	Sequence 8, Appli
25	202	3.9	872	10	US-09-843-676-54	Sequence 54, Appli
26	202	3.9	872	10	US-09-766-253-8	Sequence 8, Appli
27	202	3.9	872	10	US-09-766-253-54	Sequence 54, Appli
28	202	3.9	872	11	US-09-438-486-8	Sequence 8, Appli
29	202	3.9	872	11	US-09-438-486-54	Sequence 54, Appli
30	202	3.9	872	15	US-10-053-758-8	Sequence 8, Appli
31	202	3.9	872	15	US-10-053-758-54	Sequence 54, Appli
32	202	3.9	872	15	US-10-054-295-8	Sequence 8, Appli
33	202	3.9	872	15	US-10-054-295-54	Sequence 54, Appli
34	202	3.9	872	15	US-10-054-611-8	Sequence 8, Appli
35	202	3.9	872	15	US-10-054-611-54	Sequence 54, Appli
36	201.5	3.9	1086	10	US-09-924-154-15	Sequence 15, Appl
37	200.5	3.9	3158	9	US-09-815-242-12611	Sequence 12611, A
38	199.5	3.8	2871	15	US-10-146-473-41	Sequence 41, Appl
39	199	3.8	3899	15	US-10-171-311-4	Sequence 4, Appli
40	199	3.8	3917	15	US-10-171-311-8	Sequence 8, Appli
41	198.5	3.8	1978	12	US-10-094-466-64	Sequence 64, Appl
42	197.5	3.8	2119	12	US-09-769-744A-28	Sequence 28, Appl
43	197	3.8	1579	10	US-09-801-368-368	Sequence 368, App
44	197	3.8	3907	15	US-10-171-311-2	Sequence 2, Appli
45	197	3.8	3925	15	US-10-171-311-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-011-366-10
; Sequence 10, Application US/10011366
; Publication No. US20030054493A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-NO. US20030054493A1-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-011-366-10

Query Match 78.7%; Score 4086; DB 15; Length 2366;
Best Local Similarity 77.5%; Pred. No. 2.5e-248;
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;

QY 1 MLVYKAOQLQKVVYKFRQEDYVAILNALBEYHNMSSESVVYKYLKLDINLNTDYL 60
DB 1 MSLVNRKQLEKMANVFRQEDYVAILDALBEYHNMSNTVVYKYLKLDINSLTDIYI 60
QY 61 NTYKSGRNKALKKFKXYLTMEVLELKNLSLTPVEKNLHFIWIGQINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKXYLTVEVLELKNLNPVEKNLHFIWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKFDYDSNAFLINTLTKTIIVESATNTLTESFRENLDNPEFDYKFKRMELIY 180
DB 121 VNSDYNVNVFYDSNAFLINTLTKTVVESAINDTLTESFRENLDNPEFDYKFKRMELIY 180
QY 181 DKQKHIDYKSOIENPFIIDNIITKTVLSNEYSKDLALNKYIIESLNKTIANNRDI 240
DB 181 DKQKNFINTYKAQREENPELLIDDIIVKTVLSNEYSKEIDELNTYIIESLNKTIQNSGNDV 240
QY 241 RNLEKPADDLVLYNQELVERNLAAASDILIRISMLKEDGGVLYDVLPGIQDLPKS 300
DB 241 RNFEEFNAGESNLYEQELVERNLAAASDILIRISALKIGGYLDVDMPLGIQDLPFS 300
QY 301 INKPSDITNTSMBIKLEAIMKYKEYIPGYTSKNFDMLEEVQSPESALSSKSKSEIF 360
DB 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPEYTSSEHFDMLDEEVQSPESVSLASKSKSEIF 360
QY 361 LPDLDIKVPLEVIAFANNSVINQALISLKSYSCLDVINQIKNYKILNDNLPSINE 420
DB 361 SSLGDMESPLEVIAFANSKGIINQGLISVKDSYCSNLIVKQIENRYKILNNSLNPASE 420
QY 421 GTDFNTMTKIFSDKLASINEDNMFMKITYNLYKVGPAQVRSNTINLSGPGVYTGAYOD 480
DB 421 DNDFTNTTTFDSINAEANADGRFMELGKYLVRGFPDVKTIINLSGPEAYAAAYOD 480
QY 481 LLMFKDNSTNIHLLEPELNFEPKTKISQTEQETISLWSFNQAPAKSQFEYKKGYPE 540
DB 481 LLMFKEGSMNHLLEADLNFNFESKTNISOSTEQEMASLWSFDDARAKAQFEYKRNYPE 540
QY 541 GAGEDNDNDFAGNTVLDKDYYSKILSSMKYTKNEYIHYIYVQLQDKISYASCNLFPSK 600
DB 541 GSLGEDNDNDFSQNIYVDKEYLLEKISSLARSSEYIHYIYVQLQDKISYEAACNLF 600
QY 601 DPVSSILYOKNTGSETAYVYVADAEIKEIDKIRIPYOISNKRNIKLTFIGHKGSEFT 660
DB 601 TPYDSVLFGKNTEDSIAYYYNPGDGEIQEIDKYKIPSIISDRPKLTFIGHKGDEFNT 660
QY 661 DTFANLDVDSLSSEIFETILNLAADISPKYIEINLLGCNMFSYSIYAETYPFKLLKTK 720
DB 661 DIPAGFDVDSLSTEIAADLAKEDISPKSIEINLLGCNMFSYSINVEETYPFKLLKVK 720
QY 721 DRUSELMPISQDSITVSANOQEVNRINBEKREILDHSGKWNKKEBSIIKDISSEYISF 780
DB 721 DKISELMPISQDSITVSANOQEVNRINSEGRRELLDHSGEWINKKEBSIIKDISSEYISF 780
QY 781 NPKENKIIVKSKYLHSLTLQEIIRNANSDDIDLEKKVMLTECEINVASNIDRQIVEGR 840

DB 781 NPKENKIIVKSKYLHSLTLQEIIRNANSDDIDLEKKVMLTECEINVASNIDRQIVEGR 840
QY 841 IEEAKNLTSDSINYIKNEFKLIESISDLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
DB 841 IEEAKNLTSDSINYIKNEFKLIESISDALCDLQKQNELEDSHFISFEDISDEGFSIRF 900
QY 901 INKETGNSIFETEKEIFSEYATHISKEISINIKDIFDNVNGKLVKKNLDAHEVNTLN 960
DB 901 INKETGESIFVETEKTIPEYANHITEISIKIGTIFDTVNGKLVKKNLDTTHEVNTLN 960
QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETD 1020
DB 961 RAFFIQSLIEYNSKESLSNLSVAMKVQVYAQLFSTGLNTITDAKVELVSTALDETD 1020

RESULT 2
US-10-011-366-6
Sequence 6, Application US/10011366
Publication No. US20030054493A1
GENERAL INFORMATION:
APPLICANT: Williams, James A.
Kink, John A.
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
DISEASE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/011,366
FILING DATE: 16-NO. US20030054493A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/957,310
FILING DATE: 23-OCT-1997
APPLICATION NUMBER: US 08/329,154
FILING DATE: 24-OCT-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6

Query Match 53.6%; Score 2783; DB 15; Length 2710;
Best Local Similarity 52.1%; Pred. No. 3.2e-166;
Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;

399	Qy	TYCISNGISNTDDISRLDFLNVJLKLSEPVQNDPKSTVEKNKRDIDLLKNTISQKFGDRF	450
239	Qy	---DIRNLEKFADELVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDILPGIQP	295
459	Db	QLRDINTLESFKKQDYFFYOQEMLLRWNYAAASDQVRINILKEYGGIYTTDILPAYSD	518
296	Qy	DLFKSINKPDSITNTWSWKIEALIMKYK--EYIPG--YTSKNFMDLDEEVQRFESGALS	351
519	Db	KVSQIINE-KSDDKRFPEDLKLRRISSEILSLINGEKYSIKH-DGLDETTLNQLNLI-	575
352	Qy	SKSDKSEIFLDDIKVSPLEVKAFANNSVI-----NQALISLSDSYC	395
576	Db	SEIEK-----LTIDDY-FKPVETKVURDTFKIKYQKWTWNWIRGNMFLMTHKSGK	630
396	Qy	SDLVINQIKNRYKILNDNLNPSINEGDTFNTMTKIFSDKLASINE-----DNM	444
631	Db	IDFILSGQKQYLEL-QRIRDNISYNMLFYIT-----EDLKSNNVAIGGIPAKKYLEHG	684
445	Qy	MFMIKITNLYKVGAPDVRSTINLSGP-----	471
685	Db	LF-----SEYRQDGTIPYVYSTNLNSGPDIMRMQMKKYKSLGRICEVHIDKNKLSDVNFL	740
472	Qy	GVYTGAYODLLMF---KDNSTNHLLEPE-----LRNFEPKTKISOLT---EQEITS	518
741	Db	GVYASSNKDNKSPWNLNPSVGVINDITPDDSSWAVRNNDINKILFEKINCHVEKULPTS	800
519	Qy	LWSFNQAPAKSQFBEYKGYFEGALGEDNDIDFAQNTVLDKDYVYSKILSSMKTRKEYI	578
801	Db	LY-----YEIDSRSPFOGW---DNKSIKHVTEINKDLI-KDINLLLTSSNIDV	844
579	Qy	HYIVLOQDKLISYASCNLFSKDPYSSILYQKNIEGSETAYYYVVAADAEIKEDYRIPY	638
845	Db	KLLIKL--DRELYAISKKI-----DNPALARSINTLQLQANTVYTSNTEPENTINFIY	896
639	Qy	QISNKRN-----IKLTFIGHGSEFNDDTEANLDVDS-----LSSSEIT	677
897	Db	DFYRKQDILLSAKL-----FSRDADTKIIVTNSVMEKVVFLREVISC	942
678	Qy	ILNLAKADISPKYIEINLLGCMFYSYIABETYPGKLLKIKORVSELMPISQDSI--	735
943	Db	VLRSKKVD---SYINEN-----KKNLSKEDA--GALRDYAKLKKMELFSLMDDDGYKK	990
736	Qy	TVSANQVEVRINEBCKREILDHSGKWINKESIIKDISSKEYISFNP-----K	783
991	Db	IITTNAY---IKERDKL-----SGIYIENSIISGHESFDIISNQHENGDLSTVQPK	1042
784	Qy	ENKIIVSKYIHELSTLLOEIRNNANSSDILEKKVMLTECEINVAS-----NIDROIVE	838
1043	Db	KFEFVKSE-LSSAKSIDDKNK-YITDPTKENVLYHQLDSDIKERIAFLDISHAYP	1100
839	Qy	GRIBEAKNLTSDSINYIKNEPKLIESI-----SDSLYDLKHQNGL	878
1101	Db	GSLLLEKLQLS---GYVPSDINIITAEYLLASGVSGHYSHGVYWPAPSDKLELLRRHTK	1156
879	Qy	DDSHPISEDISKTENGPRIRPINKETGNSI---PIETEKEIIFSEYATHISKEISNIKOT	935
1157	Db	SNSEWI---EKITP-----YVYDILSDNVSNVLRPPLSEQKKILNDIKLEISKSVS--	1207
936	Qy	IFDNNVGKLVKKVNLDAAEVNTLNSAPFIOSLIEYNTTK-----ESISNLSV---	983
1208	Db	YFMKLTEQSSVIGIKYSVDPRYRNEFLSLPYNQNLTPFMYRYFEMLYDIHIGIEN	1267
984	Qy	-AMKVQVYAQLFSTGLNTITDASKVVEL	1010
1268	Db	KANREFIYSKFSNLNLDPLINDERVNL	1295

RESULT 4

RESOL 4
US-10-304-095-6

US-10-304-095-0
: Sequence 6, Application US/10304095

Publication No. US20030134275A1

; GENERAL INFORMATION:

APPLICANT: Long, David M.

QY 1002 TDAK 1006
DB 1351 TDYK 1355

RESULT 6
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishti, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-10

Query Match 4.8%; Score 249; DB 15; Length 1639;
Best Local Similarity 20.9%; Pred. No. 5.5e-07;
Matches 257; Conservative 215; Mismatches 418; Indels 338; Gaps 69;

QY 16 KRIQDEY-VATLN-ALBEYHNMSSESVVEKVL-----KLKDNINLTNYLNTYK----- 64
DB 275 KKLYQAQYDLSYNNKQLEAAHL--ISVLEKRIDTKNENIKELLDK-INEIKNPPPA 331
QY 65 KSG-----RNKAL-----KKPKY-----LTMEVLEL-----KN----- 88
DB 332 NSGNTNTLLDKNKKTEEBEKEIETAKYIKFNIDSLFTDPLELEYLREKKNIDISAK 391
QY 89 -----NSLT-PVEKNLHFIWIGQINDTAINYINQKDVNS--DYT-----VKF 129
DB 392 VETKESTEPENYGVTYPLSYN-----DINN-ALNELNSFGDLINPFDYTKPSKN 442
QY 130 VYDSN--AFLNTLTKTI-----VESATNNTLESRENLDPEFFYKPKRMELIYD 181
DB 443 IYTDNERKKPFINEIKIEKKIES-DKKSIEDRSKSLN-----DITREYKLLNEIYD 497
QY 182 -KQKHFDY--YKSQIEENPERFIIDNI--KTVLSNEYSK-DLEALN-----KYIEE-SLAK 231
DB 498 SKFNNDIDLTNFKWMGKRYSYKVEKLTHHTTPASTYENSKHNLEKLTALKATKYMEDYSLRN 557
QY 232 ITANN-----GNDIRNLEKFADELRLYNOLVERWN-----LAAASDILRI 274
DB 558 IVEKELKYKYNKLSKIENEIETLVENIKKQEBQLEKKTIDKENPKDEKILEVSDIVKV 617
QY 275 SM-----LKEDGGVLDVILPGIQ--PDLFKSINKPDSINTNTWEMIKLEBAWK 322
DB 618 QVQKVLMMNKIDELKKTQLIKNVELKHNHVSYSYKQENKQBPY-----YLIVLKEIDK 673
QY 323 YKEVIGYTSKNFMDLDEE--VORSFESALSSKSDKSEI-----PLPLDD 365
DB 674 LKVFMPKVS-----LINEKKNIKTEGQSDNSPESTEGITGQATTPGQQAAGSALEGDS 729
QY 366 IKVS-----PLEVKI-----AFANNSVINQALISLSDYSYCSOLVINGQIKN 405
DB 730 VQAQAQEQKQAQPPVPVPEAKAQAQVTPPAVNNKTEENVSKLDYLEKLEVEPLNTSVICH 789
QY 406 KYILNDNLNPSINEGTDFTNTWKIFSDKLAISINEDNMFMKIKITNYLKVGPAPVVRST 465
DB 790 KYILVS---HSTWNE--KILKQYKITKBESKLSLSCDPLDLNFIQNNIPVMYS--MFDS 842

QY 466 INLSQPGVYTGAYQ-----DLIMFKDNSTNIHLLEPELRNFEFPKTKIS-----QLTEQ 514
DB 843 LANSLSQLFMETIYKEMVNCNLYKLDNDKIKNLLBEAKKVSTSVKTLSSSSMQPLSLTPQ 902
QY 515 EITSLSWSPNQAQAKSQPEYKGYPEGALGEDDNDLDFRQNTVLDK--DYVSKILSSMKT 572
DB 903 DKPEV-SANDDTSHSTNLNLSKLPELITSLGKNKIYQELIGQSSNFYEKILKDSOT 961
QY 573 RNKEYIHVQLQGDKISYEASCNLFSDPYSSILYQKNIEGSETAYYYYVADAIEKID 632
DB 962 FYNESFTNFVSKADDIN---SLNDESK-----RKKLE-----EDIN 995
QY 633 KYRIPIQIS---NKRNIKLTFIGHGKSEFNTDTFANLDVDSLSSEIETILMLAKADISP 688
DB 996 KLKKTQLQLSFDLYNLYKLERLPDKKTVGYKMQIKKLTLLKQLESKLSLN---NP 1052
QY 689 KYIENLIGCMNFSYIY-----ABETPG-KLLK-----TKORVSELMPP--S 729
DB 1053 KHV-----LQNFSPFNKKKEAIEAETENTLENTKILLKHYKGLVKYNGESSPLKT 1104
QY 730 ISQDSITVSAN-----QYEVRIINEBK-----REILDHSGKWINKESIIKD-ISK 775
DB 1105 LSESIQTEDNYASLENFKVLKLEGKLDNLEKKLSYLSGLHLHIAELKEVIRNK 1164
QY 776 EYISFNPKNKIIVKS-----KYLHE---LSTLQEIERNNANSSDIDLEKKVMLTEC-- 824
DB 1165 NYTGNSPSENNTDVNNALESYKFLPEGTDVATVSE--SGSDTLEQSQPKKPASTHVCA 1222
QY 825 ---EINVASNDRQI-----VEGRIBEAKNLTSDSINYIKNEPKLIESISDSLYDLKH 874
DB 1223 ESNTITTSQNVDDVDVVIIPFGESE--DYDDLQGVVVTGEAVTPSVIDNI----- 1273
QY 875 QNGLDSDHFISEPIDISKTEGPRIRFINKETGNSIFIETEKEIFSEYATHISKEISNID 934
DB 1274 -----LKIENEYEVLYLKLPLAG--VYRSLLKQLENNVMTF-----NVNWKD 1313
QY 935 TI-----FDNY-NGKLVKKNLDAHAHEV-----NTLSAPFIQSILIEYNTTKESL- 978
DB 1314 ILNSRFNKRNFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSID 1373
QY 979 SNLSVAMQVYVYQAQLFSTGLNTITDASK 1006
DB 1374 TDINEFANDVLGYKILSEKYSKSDLSIK 1401

RESULT 7
US-10-032-585-7646
; Sequence 7646, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7646
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1881)..(1881)
; OTHER INFORMATION: X-any amino acid
US-10-032-585-7646

Query Match 4.7%; Score 246; DB 12; Length 1881;
Best Local Similarity 19.9%; Pred. No. 1e-06;
Matches 234; Conservative 187; Mismatches 360; Indels 396; Gaps 56;


```
QY 343 QRSFESALSSKSDSEIFLPLDDIKVSPLEVKIAPANNVINOALISLSDSYCSDLVINQ 402
Db 556 DQIIQURL-----EVF-----NKEVNTKLTATWINNGIILKSDHLSKLTQFL 599
QY 403 IKRUYKILNDNLNPSINEGDFNTTMMKIFSDKLASISNEDNMFMKIITNYLKVGFPADV 462
Db 600 IQNEQEGDDNEEDS--STNTRLSVLQSPKQVQYRKRKNPSSWFLNMIISGVTTKSFMVV 658
QY 463 RSTINISGPGVYTAQVQDLMFKDNTNHL-----LEBELRNFEP 504
Db 659 ARKISISGD--WKIOAQAIIKFGDNAQOELSFPESKVPDKPKRTVDCSIESQLNFEFF 716
QY 505 KTKISOLTEGIEISLWSFNQARAKS-----QPEYKKGV-----FECA 542
Db 717 KT--LSTLSSNLILDHMILBELQVRSIKFKLDWETIDDNKIFSKFQONHDSIKRKNSH 775
QY 543 LGEDDNLDAFQNTVLDKDYVSKKILSSMKTRKEYIHYIVOLQDQKISYEASCNLFSKDP 602
Db 776 INIDVAVDVADNT-----STTNTKYI-----RAPKLKYE--SMF--- 808
QY 603 YSILYQKN---TEGSETAYVYVADAEIKEDKYRIPYQISNKRNIKLTFFIGHGSEFN 659
Db 809 ---LIYNDKLLPTYN SATILFLKI---ELVESKNMYLKL-FGNLRNLQI---KNTSEDI 857
QY 660 TDTFANLVDLSSESIEITILNLAADISPKYIEINL----- 695
Db 858 QKULHMIDEANQYFEDNTVDLSTVINEPKTLNLI FNTLNKLSLILKLDOLNKSNTV 917
QY 696 -LGCNNFYSIYAETYPGK---LLKIKDRVSELMPISODSITVSANQYEVRIINEEG 750
Db 918 VLDSMDNITINIKDYNDNDKLIILIKLPEQATDSIQLMKNGSTTTTDEIDLRNNIL 977
QY 751 KREILDHSGKWINKESIKIDISKEYISNPENKIIIVKSKYLHELSTLLQIR----- 805
Db 978 EFELI---LKFLNQ---YLRESQSNH--NRQOISIIKIIQVLTINPILQSTKAINOQ 1028
QY 806 -----NNANSDDILEKKVMLTECEINVASNIDQIVEGREAEAKNLTSDSINYKNE 858
Db 1029 LAELKRINTNNNNNAKRASIL-----KLSNGLYKLYFNLIISLTLQLV 1075
QY 859 FKLIIESDLYDLKHQGLDDSHFISFEDISKTENGFRIRFINKETGNSIFITEKEIF 918
Db 1076 FFMNSTGSLNKKIQDKIMINLSLIKDFDRFSKPNGNQ---DSTNGHLIKISPKDSLI 1131
QY 919 SEVATHISKEISNIKOTIFDVGKLVKKNVLDAAHEVNTLN 960
Db 1132 NE-----NLKFKNLFEIIFKNINELLTK-----SKQIKSLN 1163

RESULT 9
US-09-820-843A-73
; Sequence 73, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 73
; LENGTH: 1805
; TYPE: PRT
; ORGANISM: M. genitalium
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|1045905
US-09-820-843A-73
, Query Match 4.3%, Score 221, DB 11, Length 1805;
```

```
Best Local Similarity 18.7%; Pred. No. 3.6e-05;
Matches 195; Conservative 171; Mismatches 367; Indels 308; Gaps 51;
QY 87 KNSLTPEKLNHLPTWIGQINDTAININQWQVNSDYTVKFVYDSNAFLINTLKTTIV 146
Db 29 KNRSTVDK-----IENQLLKEIKSLEDE-----LKNLKLGN 62
QY 147 ESATNTNTLESFRENLDPEFDVNFYRKRMEIYDKQKHFD-----YKSOI-- 194
Db 63 QADNPEDL---KKINHLEVDLNLRLVNEYKQFOK-NHWVDKVSSELNLTFRFYKNELTR 118
QY 195 ---EENPEFIIDNIIKTYLSN---EYSKOLEALNKYIE---ESLNKITAN-NGNDIRNLEK 245
Db 119 LQENADFL--NSKYANLANFQYVHNKLNDFHRLIENQNTINRLNQINGN---QN 171
QY 246 FADELVLRYNQL-VERKNLAAASDIILRISMLKEDGGVYLDVDILPGIQ-----PDL 297
Db 172 LIDNNVALLQNPITVEKNY-----LLNVLDOLYNELDQLENQKRLLSIEYENT 221
QY 298 FKSINKPDSITNTSWMIKLEAIMKYKEYIPGYTSKNFMDLDEEVORS-----PESALS 351
Db 222 YRELYSADNELQNVYENIDQNOI-QFKHQYTYRDE-LSQLERKQLTKQELVDKESALR 279
QY 352 SKSDKSEIFL-----PLDDI-----KVSPLVETKIAFANN-----SVINQALI 388
Db 280 VKIDDADFYINARLAELDDVAKOLSFQDGIITKONAHQVEDKLVALNKEKDRLNTQKEAFF 339
QY 389 SLKSDYCSDLVINQIKNRYKILNDNLNPSINEGTDFNTMKIFSCLKLASISNEDNMFM 448
Db 340 NLRSQALID--INKLQOENELFAHLEHQONB---FEQKQ---SDSLKLETE----- 384
QY 449 KITNYLVKGFADPVRSTINLSGPGVYTGAYQDLMFKONST--NIHLLEPELNRNFEPPK 506
Db 385 -----YKALOHKINEFKNESATKSSELLNOERLEFEKRE 419
QY 507 KISOLTEOITSLWSFNQARAKSQPEYK---KGYFEGALGEDDNLDPQAQNTVLDKQYV 562
Db 420 IDTLTQASL---EYEHQRESSQLLKQNEVKHQFHQ-----NLEVAKKE-LDKE-- 465
QY 563 SKKILSSMKTRNKEVIH-----YIV-----OLOQDKISYASCNLFSK 600
Db 466 -BNLLDQOQKVDSEAI FOLKEKVAQERKELELYLVKKQKQDKQKQENELLFFEKQLKHQA 524
QY 601 DPYSSILYQKNEGSETAYV---YVVAADAEIKEDKYRIPYQISNKRNIKLTFFIGHGKE 657
Db 525 D-----FENELEAKQOELFEAKHALERSFIKLEK-----EKDLNTKAQOIANEFSQ 571
QY 658 FNTDTFANLDVD--SLSSIEITILNLAADISPKYIEINLLGCMNFSYSIYABETY----- 711
Db 572 LKTDKSKSADPELMLQNEVENL-----QOEKQKLFQERTYPERNA 611
QY 712 -----PGKLLIKIKDRVSELMPISODSITVSANQYEVRIINEEGKEIL-----DH 757
Db 612 AVLSNRLQOQREBELLOQKTELDTOLTKSPFEQERLINOQREHKELVASVEKQKELGKQLQDF 671
QY 758 SKKWINKESIIKDISKEY-ISFNPENKNIIVKSKYLHELSTLLQEIERNNANSDDILE 816
Db 672 SQTSLNAS---KNLAEREMAKFKEKEIEATEKO-----LLNDV--NNAEVIQADL- 717
QY 817 KKVMLTECEINVASNIDQIVEGREIEEAKNLTSDSINYIKNEFKLIESISDLYDLKHQN 876
Db 718 -----AQLNQSLNQERSELQNAQRIADPHNDLSKKL-NEVEL--SLQKRLQELQTL 767
QY 877 GLDDSHFISFEDISKTEGFRIRFINKETGNSIFITEKEIFESEVATHISKEISNIKDTI 936
Db 768 ANQKQH--SYONQAYFE-----GELDKLNREKQAFNLNRKKQTMEDVDAIKQRL 813
QY 937 EDNVNGKLVKYNLDAAHEVNTLNSAFFIQSLIEYNTTKEISLNSLVAMKVQVYAQLFST 996
Db 814 SKHQALNMQQAELD--RKTHELNRAF-----LNHDADQKSLQD----- 850
QY 997 GLNTITDASKVVELVSTALDE 1017
```


Qy	389	SLKDSYCSDLVIN--QIK-----NRVKI-----LMDNLNPSINEGDTNMTTKIF-----SD	433
Db	766	HENKEYNSSYKINSEQIKRQCHQFKIDCNVEQRDDNNVVTKV--GTTKDCQCKFKNVTQD	824
Qy	434	KL-----ASISNEDNMFMKIT---NYLKVGAPDVRSTRINLSGGCVYTGAY	478
Db	825	RLKEGQERVIKVEAKIKNDMMVQBOQKTEKHKMDVQFIEB--KDNVOHINVQMDMV	882
Qy	479	QDLILPKDNNSTNIHLLEPRLNPEFPKTKISQLTEQETSLWSFNQARAKSQFEEYKKG	538
Db	883	QDM-----DVQDINVQMDVQNVINQDINQDMDVQNNISITLKS--TSCQDSESR---	933
Qy	539	PEGALGEDDNLDAQNTVLDKDYVSKILSSMKTRNKEYTHYVQLQGDKISYEASCNLF	598
Db	934	--DAPGGDQNESLDEKDSMEKSEKSKKKKGSKRKNKD--TNLTLKSDSIQ--KSKTTLD	987
Qy	599	SKDPYSSILQKNLEGSETAYYYV--ADABIK-----IDKRPYQISNKNKIKLFI--	651
Db	988	DKRNVVVTPIGHINHGKTSFLDYICKTNEQKKEYGLITONIRAFKATVRNNFTFLAVDT	1047
Qy	652	-GHCK-----SEPTD-----TPANL	666
Db	1048	PGHEAFPMRSGVKISDLISLVISGDEGIEQETVECIKUEKFIKIIITAIKTVDPNV	1107
Qy	667	DVDSL-----SSEIETI-----LNL-----	681
Db	1108	DVDRIINDLLYHDITTELNGEIQVVECSIYKEESIDKLLDAIYLESEFLNLQTNPKKH	1167
Qy	682	--AKADISPKYIE-----INLL--GNMPSYSIYABETPGKLLKIKORVELMPS--	729
Db	1168	EQAGVVLDSYIDKNGIVSINLLQNGVLNINDHFYTGSSY--GKVKI--LKDHLNKNIKSAY	1225
Qy	730	-----TSQDSITYSANOYVRINEEGKREILDHSGKWINKESIIKDISSEKYEISF	780
Db	1226	PSDPIKIIGNKNSVPVAGDKFYVENEALAKEIAEH-----NKNKMLTWEINNFTYDQT	1280
Qy	781	N-----PKNKIIVKSKYLHE-----	796
Db	1281	NNNRYKDFIISRENKIGSGGILGENNLKNDIDGNMTRDDNMTDDNMTSDDNMTDCNR	1340
Qy	797	-----LSTLQER-----NNANSSDIDLEKKVMLTEC	824
Db	1341	TNNDNITSDDNMSNDYDKIKETKMYTNKNSFKDDFLKIHLLNTNENVINMDPSTHIGKN	1400
Qy	825	EI-----NVASNIDRQIVEGRIEAKUL-----TSDSINYIKQB--FKLTESISDSLYDL	872
Db	1401	EIKTIYNYIYIKCDKQ--GSIEVLKNCMLKLQKEDSICKKNKIIVADIIGNVTSS--DI	1455
Qy	873	KHONGLDDSHFISP-----EDI-----SKTENGRFIRFINKETGNSFIFETEREIPS	919
Db	1456	KYATSP--NATIIAPGVKLSNDIKGSKNSGSKNNNNPIIYSN-----VLYELIENVEK	1508
Qy	920	EYATHISKE--ISNIKDT-----IPD-----NVNGKLKVKYNLDAAEHVNTL--NSAFFIQ	966
Db	1509	EMEKLSKPKMGLKGTAQILKVPENISKLGVAGCIVKGTISINSIRILRNDKVIYMG	1568
Qy	967	SLIEYNTKSGLSNLSVA	984
Db	1569	KIISIKIVKEKTQVTEA	1586

RESULT 13

```

RES001 13
US-09-815-242-5809
; Sequence 5809, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

```

```

Db      607 RIRYKNVILKPAWKINEMVPLPKNEWDQFLKYQ-----EQFNIPNIVNLV-- 655
QY      559 KDYVSKKILSSMKTRNKEYI-----HYIVLOQDKISYASCNLFSDOPY 603
Db      656 --YGDNKLILLNLNLANHRYLLMKEYKKHKKRVLVESFLPQSKNDHV-YEIVTPIYKSSY 712
QY      604 SSILYQKNIEGSETAYYYVADAIEIKDYRIPYQISNKRNIKLTFIGHGSEFNWTF 663
Db      713 -----RGPEI-----EIPKYK-----NTDIEYDKDWF 734
QY      664 A-NLDVDSLSEIETILNLAADISPKYIEINLLGCNMFPSYSIYAEETPGKLLKIKDR 722
Db      735 ALHHIDKPSQDTFIIDNL-----YFVKHLKDKGD 765
QY      723 VSE--LMPISQDSITVSANQYEVRI--NEEGKREILDHSGKWINKESIIKDSSKEYI 778
Db      766 IDQYFLMRYIKQGD-----LKLRLFRNDENVAEYISILKNWLPVHRQ-TTEVSDYEFV 818
QY      779 SFPNKENKIIVKSKYLHELSTLLQEIEN---NANSSDIDLEKKVWLTEC-----EINVAS 830
Db      819 SYEPFPRYGGKNT-INEIEAFEDYTNLAVNIENDFPKDFRPIYVAISIMYLFEMFIS 877
QY      831 NIDR-QIVEGRIEAKNLTSDSINYIKNEFKLIESDSL-YDLKHQNGL-----DDSHF 883
Db      878 NEERMEIVNNYV--PTSFKSKDIRPFPKNELVTICNPANNFEYMAKHYSGIYRILKDGNOI 935
QY      884 IS--FEDISKTEGFRIRFINK---ETGNSIF-IETEKEIFSEYATHISKEI 929
Db      936 LSKLNEGLKLTTKRSRIIGSLIHMCRNRFQIDKQDTF---VLSIVKEI 984

```

RESULT 14

```

US-09-815-242-12899
; Sequence 12899, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Hasebeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12899
; LENGTH: 1002
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-12899
; Query Match          4.0%; Score 209.5; DB 9; Length 1002;

```

```

Best Local Similarity 19.2%; Pred. No. 8.5e-05;
Matches 206; Conservative 159; Mismatches 354; Indels 353; Gaps 56;

QY      9  LQKMYVVKRIQEDSY-----VAILNALBEYHN-----MSBSSVVEKYLKLDINNLTDN 58
Db      120 IQK--YVKV--DGEWLKLVSYIESIDBYQNLKVIWNSKAHIINDRIYL---NEQSAI 171
QY      59  YLNTYKSGRNKALKPKPEYLTMEVLELKNLSLTPEKNLHFEIWIWGQINDTAIINYQW 118
Db      172 YLNNNKOTS-----FSIKNSL-----LVFIKTTVTNNNITFSNLA 207
QY      119 KDVSNDYTKVFDYDVSNAFLINTLKXTIVESATNNTLTSPRENLDNPEPDYNK-----F 171
Db      208 EKINGEPEINDITKVYVIHNLVSKIEIYSTIRPPL-SYSDNLN---YLNLKLSLHNDFF 263
QY      172 YKRMEEIYYDKQKFIIDYKYSQIEENPEFIIDNI-----IKTYSNEYSKDLALNK 223
Db      264 VKKIREI---QKLILAVEKTEIGFGEELYKDIHHMKALPKCKNYL--QIDTKIDMINN 317
QY      224 YIEESLNKITANGNDIRNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGV 283
Db      318 YLHQDI-----ATNISEA-----YLLMLLGRNNIG----- 343
QY      284 YLDVDILPGIQDLPFSINKPDSITNTSWEMIKLEAIMKYKEYIIPGYTSKNFMDLDEE-- 341
Db      344 FTDLKVL-----HNRPIEKYGFQELVNIKDLISDITGFGTTIQEET 386
QY      342 -----VORSPESSALSKSKSEIFLDDIKVSPLEVKIAPANNVINQALISUKOSY 394
Db      387 DGNNTVWLKQKFLHALRN-----DEIVINEKQVE-SLINDNEINHYPMSADV 435
QY      395 CSDIVINQIKRYK--ILNDNLNPSINEGTDTNTMKIF-SOKLASISNE----- 441
Db      436 YAEYLGRFYNOYNELIVISPLTASFAGATGRFHLLIDTETLAKLEHEKHGYQKMC 495
QY      442 DNMFMWIKITNY-----LKVGFAPDVRTINLSGGVVTGA----- 477
Db      496 DDNVEIMISINNI PKYPRHNHVNLTNDSYEYSNLG--SSNSYSKYELTLLDIIYVGATFNKL 554
QY      478 --YQDLL--MPKONS-----TNIIHLEPELRNFEPKTK-ISQLTEQBISL----- 519
Db      555 YLYSQLNKRVLFESNNMNFLECNLYL---LREISMESVKCIEPMNDVSDISPSYSP 611
QY      520 -----WSFNQ-----ARAKSOFEPYKKGYPFEGALGEDDNLQFAQNTVLD 558
Db      612 RIRYKNVILKPAWKINEMVPLPKNEWDQFLKYQ-----EQFNIPNIVNLV-- 660
QY      559 KDYVSKKILSSMKTRNKEYI-----HYIVLOQDKISYASCNLFSDOPY 603
Db      661 --YGDNKLILLNLNLANHRYLLMKEYKKHKKRVLVESFLPQSKNDHV-YEIVTPIYKSSY 717
QY      604 SSILYQKNIEGSETAYYYVADAIEIKDYRIPYQISNKRNIKLTFIGHGSEFNWTF 663
Db      718 -----RGPEI-----EIPKYK-----NTDIEYDKDWF 739
QY      664 A-NLDVDSLSEIETILNLAADISPKYIEINLLGCNMFPSYSIYAEETPGKLLKIKDR 722
Db      740 ALHHIDKPSQDTFIIDNL-----YFVKHLKDKGD 770
QY      723 VSE--LMPISQDSITVSANQYEVRI--NEEGKREILDHSGKWINKESIIKDSSKEYI 778
Db      771 IDQYFLMRYIKQGD-----LKLRLFRNDENVAEYISILKNWLPVHRQ-TTEVSDYEFV 823
QY      779 SFPNKENKIIVKSKYLHELSTLLQEIEN---NANSSDIDLEKKVWLTEC-----EINVAS 830
Db      824 SYEPFPRYGGKNT-INEIEAFEDYTNLAVNIENDFPKDFRPIYVAISIMYLFEMFIS 882
QY      831 NIDR-QIVEGRIEAKNLTSDSINYIKNEFKLIESDSL-YDLKHQNGL-----DDSHF 883
Db      883 NEERMEIVNNYV--PTSFKSKDIRPFPKNELVTICNPANNFEYMAKHYSGIYRILKDGNOI 940
QY      884 IS--FEDISKTEGFRIRFINK---ETGNSIF-IETEKEIFSEYATHISKEI 929

```

```

387 DGNIVMLKQFELHALRNN-----DEIVINEKQVE-SLINDNEINHYPHAPSADV 435
395 CSDLVINOIKRNYK--ILNDNLNFSINEGTDFTNTMKIF-SDKLASISNE-----441
436 YAEYLGRFYNOYNELIVISPLTASFNAGATFGRPHLLIDTETLAKLEKHGYQKMIC 495
442 DNMFMELIKITNY-----LKYGAPDVRSTINLSGPGVVTGA-----477
496 DDNVEMISINNIPIKPRNHNVLTNHDSEYSLNLC-SSNSYSKYELTLDIIVGATFNKL 554
478 --YQDLL----MFKDONS-----TNIHLEPELRFNFPFKTK-ISQLTEQETSLS- 519
555 YLYSSQLMKRVLFESNNMYNFLKECNLYRL---LREISMESVKCIEPMNDVSDISFSYP 611
520 -----KSPNQ-----ARAKQFSEYKGYFGALGEDNDNLDPAQNTVLD 558
612 RIRYKNWILKPAYKINEMVILPLKPNEEWDOQFLKYQ-----EQNIPINVLV-- 660
559 KDYVSKKILSSMKTRNKEYI-----HYIVQLQGDKISYEASCNLFSPKDPY 603
661 --YGDNKLNLNLSLHRYLLMKEYKHKRVLVESFLPQSKNDHV-YEIVTPIYKSSY 717
604 SSILYQKNIEGSETAYYYYYADAETKEIDKYRIPYQISNKRNIKLTFIGHGKSEFNTDF 663
718 -----RGPEI-----EIPKY-----NTDIEYDKDWF 739
664 A-NLDVDSLSEIETILNLAKADISPKYIEINLGCNMFSYSIYAETPGKLLLIKOR 722
740 ALHTHIDKPSQDTFIIDL-----YPFVKHLKDKGD 770
723 VSE--LMPISQDSITVSANQYEVRI---NEEGKREILDHSGKWINKESIIKDISKEYI 778
771 IDQYFLMRYIKQGD-----LKURLPRDENYAEIYSLTKNWLPHVRQ-TTEVSDYFV 823
779 SFNPKENKIIVKSKYLHELSTLLQBI RN---NANSSDIDLEKKVMMLTEC-----EINVAS 830
824 SYEPEFRYGGKNT-INIEAFFEYDITNLAVNIIENDPKFDRPYVAISIMYLFEMFSIS 882
831 NIDR-QIVEGRIBEAKNLTSDSINVIKNEFKLIESDSL-YDLKHONGL-----DDSHF 883
883 NEERMEIVNNYV--PTSPKSKDIRPPFKNELVTICNPANNFEMAKHYSIYRLKDCNQI 940
884 IS--FEDISKTENGFRINFINK---ETGNSIF-LETEKEIPFSEYATHISKEI 929
941 LSKLNEGKLTITTKRSRIIGSLIHMRCNRIFGIDKDOBT---VLISIVKEI 989

```

Search completed: November 5, 2003, 19:41:22
Job time : 31.375 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:25:47 ; Search time 9.375 Seconds
(without alignments)
5116.506 Million cell updates/sec

Title: US-09-126-816b-6_COPY_1_1020

Perfect score: 5192

Sequence: 1 MNLVNAQLQKVMYKFRIQ.....ITDASKVVELSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4086	78.7	2366	1 TOXB_CLODI	P18177 clostridium
2	2783	53.6	2710	1 TOXA_CLODI	P16154 clostridium
3	293.5	5.7	1251	1 RBP2_PLAVB	Q00799 plasmodium
4	273.5	5.3	2869	1 RBP1_PLAVB	Q00798 plasmodium
5	257.5	5.0	1162	1 BXEN_CLOBO	P46082 clostridium
6	256.5	4.9	1875	1 MLPI_YEAST	Q02455 saccharomyc
7	254	4.9	2136	1 YCF2_MARPO	P09975 marchantia
8	249	4.8	1630	1 MSPI_PLAFK	P04932 plasmodium
9	249	4.8	1639	1 MSPI_PLAFW	P04933 plasmodium
10	247	4.8	1679	1 YIO9_YEAST	P40457 saccharomyc
11	245	4.7	1701	1 MSPI_PLAFV	P13819 plasmodium
12	244	4.7	1682	1 MSPI_PLAF3	P19598 plasmodium
13	244	4.7	1701	1 MSPI_PLAFM	P08569 plasmodium
14	243	4.7	1726	1 MSPI_PLAFD	P04934 plasmodium
15	243	4.7	1726	1 MSPI_PLAFP	P50495 plasmodium
16	243	4.7	1928	1 MYS1_YEAST	P08964 saccharomyc
17	237.5	4.6	2339	1 RPCI_PLAFA	P27625 plasmodium
18	237.5	4.6	3433	1 UTRO_HUMAN	P46939 homo sapien
19	237	4.6	1005	1 RA50_METJA	Q58718 methanococc
20	237	4.6	1163	1 SBCC_CLOAB	Q97181 clostridium
21	236.5	4.6	1790	1 USOI_YEAST	P25386 saccharomyc
22	236.5	4.6	1957	1 SPOF_SCHPO	Q10411 schizosacch
23	235	4.5	1179	1 EXSB_BUCAP	Q8k9a9 buchnera ap
24	235	4.5	3210	1 CENP_HUMAN	P49454 homo sapien
25	234.5	4.5	1162	1 BXEN_CLOBU	Q06366 clostridium
26	232	4.5	1169	1 SMC_METJA	Q59037 methanococc
27	231.5	4.5	950	1 Y511_RICPR	Q92d36 rickettsia
28	231.5	4.5	1290	1 RA50_SCHPO	Q9utj8 schizosacch
29	231	4.4	1225	1 Y309_MYCGE	P47551 mycoplasma
30	228	4.4	1324	1 SMK2_SCHPO	P41004 schizosacch
31	227	4.4	1658	1 YM67_YEAST	Q03661 saccharomyc
32	226	4.4	1803	1 YJL3_YEAST	P47024 saccharomyc
33	225.5	4.3	1727	1 ALM1_SCHPO	Q9utk5 schizosacch

RESULT 1
TOXB_CLODI
ID TOXB_CLODI STANDARD; PRT; 2366 AA.
AC P18177;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Toxin B.
GN TOXB OR TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90326540; PubMed=2374729;
RA Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
RT "Nucleotide sequence of Clostridium difficile toxin B gene."
RL Nucleic Acids Res. 18:4004-4004(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 1271-2366 FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=92293124; PubMed=1603068;
RA Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S., Schulze J.,
Sauerborn M.;
RT "Comparative sequence analysis of the Clostridium difficile toxins A
and B."
RL Mol. Gen. Genet. 233:260-268(1992).
CC -!- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN ENTEROTOXIN
CALLED A AND CYTOTOXIN B.

34 225 4.3 1956 1 ATX1_PLAFA Q04956 plasmodium
35 224 4.3 1018 1 YC14_METJA Q58611 methanococc
36 223.5 4.3 1228 1 YHV5_YEAST P39851 saccharomyc
37 223.5 4.3 2059 1 TEGU_HSV7J P52362 human herpe
38 223 4.3 1169 1 EXSB_BORBU O51578 borrelia bu
39 222 4.3 1102 1 RPOP_AGABT P33539 agaricus bl
40 221.5 4.3 2104 1 MYS3_SCHPO O14157 schizosacch
41 221 4.3 1805 1 HMW2_MYCGE P47460 mycoplasma
42 220.5 4.2 1196 1 BXCN_CLOBO P46081 clostridium
43 219.5 4.2 4725 1 DYHC_DICDI P34036 dictyosteli
44 218 4.2 2116 1 MYS2_DICDI P08799 dictyosteli
45 216.5 4.2 1208 1 PCP1_SCHPO Q92351 schizosacch

ALIGNMENTS

Query Match		78.7%; Score 4086; DB 1; Length 2366;
Best Local Similarity		77.5%; Pred. No. 9.1e-150;
Matches	790; Conservative	105; Mismatches 125; Indels 0; Gaps 0;
QY	1	MNLVNAQLOKMYVFRIOEDYVAILNALEYHNMSESSVVEKYKLKDINNLTNYL 60
DB	1	MSLVNRKQLEKMANVFRTOEDYVAILDALEYHNMSESSVVEKYKLKDINNLTNYL 60
QY	61	NTYKSGRKNALKKFKEYLTMEVLEKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120
DB	61	DTYKSGRKNALKKFKEYLTMEVLEKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120
QY	121	VNSDYTVKFDYDSNAFLINTLKTIVESATNTNLTLESFRENLDNPEFDYNNFKRMEIY 180
DB	121	VNSDYNNVFDYDSNAFLINTLKTIVESAINDTLESFRENLDNPEFDYNNFKRMEIY 180
QY	181	DKQKHFIDYKSIENPEFIIDNIIKTYSNEYSKDLALNKYIEESLNKITTANNNDI 240
DB	181	DKQKNFINYKQREENPELLIIDIVKTVLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240
QY	241	RULEKFADEDLVRLNQVELVERNLAAASDILRISMLKEDGGYLDVDILPGIQDLPFS 300
DB	241	RNFEEFKNGESFNLYEQELVERNLAAASDILRISALKEIGGYLDVDMLPGIQDLPFS 300
QY	301	INKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVORSPEALSCKSKSEIF 360
DB	301	IEKPSVTVDWEMTKLEAIMKYKEYIPYTSSEHFDMLDEEVQSSVESVLASKSKSEIF 360
QY	361	LPDDIKVPLEVKIAPANNVINQALISLKOSYCSDLVINQIKNRYKILNDNLPSINE 420
DB	361	SSLGDMEASPLEVKIAFNKGIINOGLISVKOSYCSNLIVKQIENRYKILNNSLNPASE 420
QY	421	GTFDNTMTKIFSKLASINEDNMFMWIKITNLYKGFAPDYRSTINLSGPGVYTGAYOD 480
DB	421	DNDFTNTTTFDISIMAEANADGRFMELGKYLVRGFPDVKVTTINLSGPAYAAAYOD 480
QY	481	LMFMKDNSTNIHLPELARNFEPKTKISOLTEQETISLWSFNQARAKQFEYKKGYPE 540
DB	481	LMFMKSGSNHILIEADLRNLFELSKTNIQSOTQEWASLWSFDDARAKQFEYKKNYFE 540
QY	541	GALGEDNDLDFQNTVLDKDYVSKKILSSMKTKRNEYIHYIYVOLQGDKIISYASCNLPSK 600
DB	541	GSLGEDNDLDFSQNVVDKEYLEKISSLARSSERGYIHYIYVOLQGDKISYEACNLPAK 600
QY	601	DPVSSILYKQNIETGSETAYYYVADAEIKEDYRIPYQISKNRKNILKTFIGHGSEFNT 660
DB	601	TPYDSVLFOKNIBDSIAYYNPGDGEIQEIDKYKIPSIISDRPKILKTFIGHGKDEFNT 660
QY	661	DTFANLDVDSLSSEIETILNLAKADISPKYIEINLLGCNNFYSIYAEETYPGKLLKIK 720
DB	661	DIFAGFDVDSLSLIEAIDLAKEDISPKYIEINLLGCNNFYSINVEETYPGKLLLVK 720
QY	721	DRVSELMPSISQDSITVSANQYEVRIINEGKRILDHSGKWINKESIIKDISKEYISF 780
DB	721	DKISELMPSISQDSIIIVSANQYEVRIINSEGRRLDHSGEWINKESIIKDISKEYISF 780
QY	781	NPKENKIIKSKYLHELSTLLOIRNANSDDIDLEKKYMLTECEINVASNIDRQIVEGR 840
DB	781	NPKENKITVKSKNLPELSTLLOIRNANSDDIDLEKKYMLTECEINVSINIDTQIVEER 840
QY	841	IEEAKNLTSDSINYIKNEFKLIESISDSYLDLKHQGLDDSHFISPEDISKTENGPRIF 900
DB	841	IEEAKNLTSDSINYIKDEFKLTIESISDALCDLKKQNELEDSHFISPEDISDEGFSIRP 900
QY	901	INKETGNSIFIEPEKIFSEYATHISKEISNIKOTITPDNVNGKLVKKNVLDAAHVNTLN 960
DB	901	INKETGESIFVETEKIFSEYANHITEEISKIGTIFDTVNGKLVKKNVLDTHTEVNTLN 960
QY	961	SAFFIOSLLEYNTKESLNSLVAMKVQVYAOQLFSTGLNTITDASKVVELSTALDETID 1020
DB	961	AAFFIOSLLEYNSSKESLNSLVAMKVQVYAOQLFSTGLNTITDAAKVVELSTALDETID 1020

RESULT 2	
TOXA_CLODI	STANDARD; PRT; 2710 AA.
ID	TOXA_CLODI
AC	P16154;
DT	01-APR-1990 (Rel. 14, Created)
DT	01-FEB-1996 (Rel. 33, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	Toxin A
GN	TOXA OR TCDA.
OS	Clostridium difficile.
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC	Clostridium.
OX	NCBI TaxID=1496;
RN	[1]_TaxID=1496;
RP	SEQUENCE FROM N.A.
RC	STRAIN=VPI 10463;
RX	MEDLINE=90221894; PubMed=2109310;
RA	Sauerborn M., von Eichel-Streiber C.;
RT	"Nucleotide sequence of Clostridium difficile toxin A.;"
RL	Nucleic Acids Res. 18:1629-1630(1990).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=VPI 10463;
RX	MEDLINE=90129305; PubMed=2105276;
RA	Dove C.H., Wang S.Z., Price S.B., Phepals C.J., Lyster D.M.,
RA	Wilkins T.W., Johnson J.L.;
RT	"Molecular characterization of the Clostridium difficile toxin A
RT	gene.;"
RL	Infect. Immun. 58:480-488(1990).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=VPI 10463;
RA	von Eichel-Streiber C.;
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC	-!- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
CC	REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
CC	DIFFERENT OLIGOPEPTIDES.
CC	-!- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
CC	ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL
CC	DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE
CC	CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
EMBL; X51797; CAA36094.1; -	
EMBL; M30307; AAA23283.1; -	
EMBL; X92982; CAA63564.1; -	
InterPro; IPR002479; CW binding.	
Pfam; PF01473; CW binding_1; 28.	
Pfam; PF04488; Gly_transf_sug; 1.	
Toxin; Enterotoxin.	
SEQUENCE 2710 AA; 308052 MW; 0A6B52CE84C14421 CRC64;	
Query Match	
Best Local Similarity 53.6%; Score 2783; DB 1; Length 2710;	
Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;	
QY	1 MNLVNAQLOKMYVFRIOEDYVAILNALEYHNMSESSVVEKYKLKDINNLTNYL 60
DB	1 MSLVNRKQLEKMANVFRTOEDYVAILDALEYHNMSESSVVEKYKLKDINNLTNYL 59
QY	61 NTYKSGRKNALKKFKEYLTMEVLEKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120
DB	60 NRYKTSRRNALSNLKDILKEVILKNSNTSPVEKNLHFWIGVEVDIALEYKQWAD 119
QY	121 VNSDYTVKFDYDSNAFLINTLKTIVESATNTNLTLESFRENLDNPEFDYNNFKRMEIY 180

Db 120 INAEYNIKLWYDSEAFVNLTKAIVESSTTEALQLEEEIQNPQPDNNKPKYKCRMEFTY 179
 QY 181 DKQKHPIDYKQIEENPEFIIDNIITKYLSEYKDLKALNKYIEESLNKTIANNNDI 240
 Db 180 DRQKRPINYSKQINQKPTPTDDIIKSHLSEYRDETVLESYRTNLSRKINSHGIDI 239
 QY 241 RNLEKPADLRLVNLQELVLRWNLAAASDIIRISMLKEDGGVYLDVILPGIQDPLPKS 300
 Db 240 RANSLPTEQELLNIYSQELNLRGNLAAASDIVRLALKNFYGVYLDVDMPLGHSPLPKT 299
 QY 301 INKPSITNTSEMIKLEAIMKYKEYIPGYSKPNFMDLDEEVQSPESALSXSXSEIF 360
 Db 300 ISRPSIGDRWEMIKLEAIMKYKYINNTSENFDPKLDQQLKONPKLJIESKSEIF 359
 QY 361 LPLDDIKVSPLEKIAFANNVINQALISLKDSYCSDLVINQIKRYKILNDLNPISNE 420
 Db 360 SKLENLNSDLAIKIAFALGVSINQALISKQSYLTNLVIEQVKRYQFLNHLNPAIES 419
 QY 421 GTDFNTMTKIFSDKLASISNEDNMFMKITYNLKYGFAPDVRSTINLSGPGYTTGAYOD 480
 Db 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQGFMPPEARSTISLSGPGAYASAYD 479
 QY 481 LLMFKONSTNIHLLPELNFPPPKTKISQLETEITSLWSFNQAPAKSQFBEYKGYE 540
 Db 480 FINLQENTIEKTLKASDLIEFKPPENNLSQLTEQIEINLSWSPDQASAKYQFEKYRDTG 539
 QY 541 GALGEDNDLPAQNTVLDKDY-VSKKILSS--MKTENKEYIHYVLOQDKISYASCNL 597
 Db 540 GLSUEDNGVDFNKNTALDKNYLNKIPENNVEEAGSKNYVHYIQLQGGDDISYEATCNL 599
 QY 598 FSKDPYSSILYQKIEGSETAYYYVAD--ABIKEIDKYRIPYQISNKENIKLTFIGHCK 655
 Db 600 FSKNPKNSIIQRNM--NESAKSYPLSDGSESTLELANKYRIPERLKNKEKVKVTFIGHCK 657
 QY 656 SPFNTDTPANLVDLSLSSEIETILNLAKADISPKYIEINLGCNMFYSIYAEETYPGKL 715
 Db 658 DEFNTSEFARLSVDSLNSBISFLDTIKLDISPKNVENVLLGCNMFSDYFNVEETYPGKL 717
 QY 716 LKIKDRVSELMPISQDSITVSNAYEVRINEEGKREILDHSGKWINKKEESTIKDISK 775
 Db 718 LUSIMDKITSTPDVKNKSITIGANQYEVIRNSEGKRELLAHSGKWINKKEEAIMSLSK 777
 QY 776 EYISFNPKENKIIVSKYHELSTLQEIENNANSDDIDLEKKVMLTECEINVASNIDRQ 835
 Db 778 EYIFPDSIDNKLKAKSNIPGLASISEDIKTLDDASVSPDTKFIANNLKLNISSIGDY 837
 QY 836 IVEGRIEAKNLTSDSINYINEFKLIESISDLYDLKHONGLDSDHSFISPDISTENG 895
 Db 838 IYVEKLEPVKNIIHNSIDDLIDEFNLENVSDLEYELKKNLNDKYLISFEDISKNST 897
 QY 896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNIKOTIFDNVNGKLVKKNVLDAAHE 955
 Db 898 YSVRFINKNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVGNNLNDIQLDHTSQ 957
 QY 956 VNTLNAFTQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTAL 1015
 Db 958 VNTLNAFTQSLIDISVSNKOVNLDLSTSVKQVLAQLFSTGLNTIYDLSIQLVNLISNAV 1017
 QY 1016 DETID 1020
 Db 1018 NDTIN 1022

RESULT 3

DBP2_PLAVB STANDARD; PRT; 1251 AA.
 AC Q00759;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Reticulocyte binding protein 2 (fragment).
 GN RBP2.
 OS Plasmodium vivax (strain Belem).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN NCBI_TaxID=31273;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=16177311;
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 merozoites.";
 RL Cell 69:1213-1226(1992).
 CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 HUMAN RETICULOCYTE CELLS.
 CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC EMBL; M88098; AAA29744.1; -;
 DR Malaria; Receptor; Membrane.
 KW NON TER 1
 FT NON TER 1251
 SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 5.7%; Score 293.5; DB 1; Length 1251;
 Best Local Similarity 19.7%; Pred. No. 0.00018;
 Matches 207; Conservative 194; Mismatches 405; Indels 247; Gaps 45;
 QY 119 KDVSNDYTKFYVDYDNFLNLTKKTIVESATNNTLSEFRENLDPEFDYKPY----- 172
 Db 8 KDTSEDEKKSIEKAYEKMGNTLK-----ELEKDDKEKIEKEVEEAQIQKFIPIHDVN 63
 QY 173 -----RKRMELIYDKQHFIDYKQIEENPEFIIDNIITKYLSEYKDLKALNKYIEE 227
 Db 64 LMNDEVEKSKIVWEK---IELYKEIDEIKQ-----KT---NEY-KQGDTSNFYITE 108
 QY 228 SLANKITANNNDIRNLEKFADEDLVRNLQELVERNLAAASDILRISMLKEDGGVYLDV 287
 Db 109 QYNSATSKA-----KIQFI-----NIATTK-----KTSQTSQDI 140
 QY 288 DILPGIQDPLFKSINKPDSITNTSWEM-----IKLEAIMKYEYIPGYSKN----- 334
 Db 141 NELESIEKEVHNKQLVKQESNMEEMRKQILSMKOLLINSETIAKEISNTQALGF 200
 QY 335 -----FDMLDEEVQSPESALSXSXSEIFLPLDDIKVSPLEKIAFANNVINQA- 386
 Db 201 RENAKTKLNKTDLLQVAAIEEAKAHKNIDIALEDAQIDTEVSKIEQINREIMNKKD 260
 QY 387 -----LISLKD--SYCSDLVINQIKRYKI-----LNDNLNPSI 418
 Db 261 EIKSYLSEIKYKOKCTTEISNSKRGKQIEFKPEKNEESNKNVINEINIRNSE 320
 QY 419 NEGTFD-----NTTMKIFSDKLASISN--BDNMFMKITYNLKYGVFAPDVRSTINLS 469
 Db 321 QYVKDIEDAKQASTKVELFHKHETISNI FKEISILGVETSKQKINKAEIMKEIERH 380
 QY 470 GPGVYT--GAYQDLMFKDNSTNIHLLPELNFPPPKTKISQLETEQIEITSLWSFNQARA 527
 Db 381 NSEIQTVKGFQENLKNLNEPHYDNEDELNDKSTNAKV--LIETNLESV-----KIN 433
 QY 528 KSQFBEYKGYEGALGE---DDNLDLPAQNTVLDKDYVSKKILSSMKTENKEYIHYIQL 584
 Db 434 LSEITNIKQG-----GEKIYSKARDIMQIKATSENTAETKLEKVDQSQSYNYL---- 484
 QY 585 QGDKISYEASCNLFKSDPYVSSI-LYQKNIEGS--ETAYVYVADAE-IKEIDKYRIPYQI 640
 Db 485 --NQITTEENLVTEKNRLNGIDSTITNIEGALKESKGYEIGFLEKLEIGKRNKLVKD 542
 QY 641 SNKRNIKLTFFIGHKSEFNT-----DTFANLD-----VDSLSEIETILNLAKADI--- 696

543 ITKKSINST-VGNFSSLFNNFDLNQYDFNKNDYENKMGIEYNEPEGSINKISENLRNA 601
 687 QY -----SPKYL-----EINLG-----CNMFSYSIYAEYTPGKLLKIKDRVSELM 727
 602 SENTSYSAKTLRLAQKEKVNLLKNEEANKYLRDVKKVESF--RFIFNMKESLDKIN 659
 728 PSISQDSITVSANQYVRINEEKGRETLIDHSGKWINKESIIKDISKEYISNPKNENKI 787
 660 EMKKEQLTVNEGHNGVQVLVENIKELVDE-----NNLSDLAQATGKN-----BEIQ 708
 788 IVKSKYLHELSTLLOIRNNANSSDIDLEKKVMLTE-----CEINVASNIDRQIVEGRIBE 843
 709 ITHSTLKNRAKTLGHVDFSAKVGKIPTELALTELLGDAKLKTAQELKFESKNNVLE 768
 844 AKULTSDSINYIQVEKFLBESISD-----SLYDLKHONGLDDSHFISFEDISK-TENG----- 895
 769 TENMSKNT-----NELDVHKNIODAYKVALEILAHSDIEDTKQ-----KDSKLIEMGNQI 819
 896 -PRIRINKETGNSIFITEKEIFSEVATHISK---EISNI---KDTIFONVNGKLVKYN 949
 820 YLKVVLINQYKNKISSKSEAVSVKIGNVSKHSELSKITCSDSYDNI-----TA 872
 950 LDAAEHVNTLNSAFFIQ-----SLIEYNTTKESLNSLVAMK----- 986
 873 LEKOTELONLRSFTQETKNTWSDSKLEKIKTDFESLKNALKTLEGEVNAKASSDNHEH 932
 987 VOYQAFLSTGLNTITDASKVVELSTALDETI 1019
 933 VQSKSEPVNPALSEIEKEETDIDSLNLTALDELL 965

RESULT 4

RBPI_PLAVB STANDARD; PRT; 2869 AA.
 AC Q00788;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Reticulocyte binding protein 1 precursor.
 GN RBPI.
 OS Plasmodium vivax (strain Belem).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 merozoites";
 RL Cell 69:1213-1226(1992).
 CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 HUMAN RETICULOCYTE CELLS.
 CC -1- SUBUNIT: Homodimer (Potential).
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; M89097; AAA29743.1; -;
 DR EMBL; M89097; AAA29743.1; -;
 KW Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
 FT DOMAIN 18 2807 EXTRACELLULAR.
 FT TRANSMEM 2808 2826 POTENTIAL.
 FT DOMAIN 2827 2869 CYTOPLASMIC.
 FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
 SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Query Match 5.3%; Score 273.5; DB 1; Length 2869;
 Best Local Similarity 20.2%; Pred. No. 0.0028;
 Matches 245; Conservative 210; Mismatches 408; Indels 351; Gaps 63;
 6 KAOLQKQVVKPRIODEY-----VAIINALBEYHNMSSESVVEKYKLK-DINN- 55
 946 KALKKEIVSDSLRDKIDQYEFTEFKTSVAVTSTIOSLSKA--IDSLKRLNGSINNCK 1003
 56 ---TD-NYLTNTYKSGRNKALK-----KFKEYITMEVLELKNLSLTPVEKNLHPIWTG 104
 1004 KYNTDIDLLRSKIKTREEVQKMPKRGDKCGENTALLKSLRDKMGKINEKLN---D 1059
 105 GOIN--DT-----AINVINQKDV-----NSDYTV- 127
 1060 GRNLSLDTKKEDLLKPYSESKEIHLKXDKQKQDPQLNARIDEDIKRDVDELNVQVI 1119
 128 -----KFVYDSNAFL-----INTLKKTIVESATNNTLESFRE-----NLNDPFPDYNK 170
 1120 SENKVTLPKNSVTYIEAMHSHINTVAHCIT--SNKNEILKSVKEVEDKLNLAQEDEDYK 1178
 171 F-----YRKMEII-----YDKQKHPIDYKSOIENPEFIDNIKTYLSN-----EYSKDL 219
 1179 VKNPENKQLEAIRGSMKSKLKEVINKHVSEMTQ-----LESTANTLKSNAGKENEHDL 1233
 220 ALNKY-----LEESLNKI-----TANNGNDIRLEKFADEDLVRLNQELVERWNL 266
 1234 ELNKTGQWRDIYEKLLKIAELKEGTVNEKLANEKANKVPEPERNIIGHVLE----- 1288
 267 AASDILRISMLKEDGGVLDVDILPGIQDLPKFSINKPDSITNTSWEMIKLBAMKYKY 326
 1289 -----RITVEKDAGKV--VEEMSLKTKIEKLOE---TSDDSQNELVTTISYKLEN 1337
 327 IFGYTSKNFMDLDEEVQSFESALSCKSDSEIFLPLDDIKVSPLEVKI-----AFANNV 382
 1338 AKGY-----EDVIGKNEEDSIQLR-EKAKSLETLDKMKLVQOVNMLQSAIQGNAG 1388
 383 INQALISLKDSCDLVINQIKNRVKILNDNLNPSINGTDTNTMKIFSDKLA----- 436
 1389 ISKEL-----NELKGVIELLSTYSSILEYVKNKSSVRFQSLGANGFTK 1435
 437 SISNEDNMFMKIKITNLYKVGPAPOVVRSTINLSGPGVYTGAYQDMLLMFKDNSTNHLRP 496
 1436 ASGEKKNASARLAEKLEKQIVKDL-----DYSN-----IDD 1468
 497 ELRNFPFKTKISQITQEIBITSLMNFQARAKSQEYKGYFEGALGDDNLDPAQN--- 554
 1469 KVKKIEGKREILKMKESALT-FWBESE-----KFKQMCSSHMENAKGKKIEYLKNG 1522
 555 -----TVLDK-----DYVSKILSLSMKTRNKEYIHYIVLOQDCKISYEASC----- 595
 1523 DGGKANITDSQMEEVGNVYVK-----AEHAFHTVEAQVDKT--KAFESIVAYVT 1570
 596 ---NLFSKDPYSILYQKNIEGSTAYYYYVADAEIKEDKIRPYQIS-NKRNI----- 646
 1571 KMDNLFNESLMKEVKVCKEKKNDKAEKY---SAKLKPYDG-RIKARVSENERKISLKE 1625
 647 KUTFTGHGKSEPN-TDTFANLVDLSSEIETILNLAADISPKYIEINLLGCMPSYSI 705
 1626 KAKVEKSSQLNDVSTKSLQIDNCRQOLDVLS-----NIGRVKQNALQYFD 1674
 706 YAEETYPGKLLKIKDRVSELMPSISQDSITVSANQYVRINEEKGRETLIDHSGKWINK 765
 1675 SADKSNKSVL-----PISELGAESLDKVAKASSEYKQL-----ETVQNMESRINVE 1722
 766 ESIKIDISSKEYISFPNKENKIIKVSXYLHELSTLLOIRNNANSSDIDLEKKVMLTECE 825
 1723 EGSLTIDDKK---ITDIDNLLKMKQYEE--GLLOKIKENADKESNFE---LVGSE 1772
 826 INVASNIDROI-VEGRIBIEAKNLTSDSINY-----IKNEFKLIESISDLKHQNGL 878
 1773 INALLDPSTISIFIKLKE-YDMTGLDKNYGVKNMEIHGEF---TKSYNLIETHLSNAT 1827

```

QY 879 DSHFISFE-----DISKTENG-----PRIRFINKETGNSIFITETKEIFSEYA 922
DB 1828 DYS--VTPKQAQSLRELAKSEBEHLRRREEAIFLLNDIKKVESLKLKEMMKVSAEYE 1885
QY 923 -----THLSKEISNKTDFPNVNG-----KLVKV-----NLDAAEVN 957
DB 1886 GMRDHTSVSQLVQDMK-TIVDELKTLDNDISECSSVLNNVSVIVKVKESKHADYRRDAN 1944
QY 958 -----TNSAFF-----IQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLTGTIDTAS 1005
DB 1945 SMYESWTLNVLNPLSDEAKISSGMEFNA--EMKSNFKTDLELEIP-----SVISNSN 1994
QY 1006 KVVELVSTALDETI 1019
DB 1995 ELLKKIEQSDNDVI 2008

RESULT 5
BXEN_CLOBO STANDARD; PRT; 1162 AA.
AC 046082;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type E, nontoxic component.
GN ENT-120.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=93195515; PubMed=8450310;
RX FUJII N., Kimura K., Yokosawa N., Yashiki T., Tezuka K., Oguma K.;
RA "The complete nucleotide sequence of the gene encoding the nontoxic
RT component of Clostridium botulinum type E progenitor toxin."
RL J. Gen. Microbiol. 139:79-86(1993).
CC -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D12697; BAA02194.1; -
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
KW Neurotoxin.
SQ SEQUENCE 1162 AA; 136856 MW; 96468DEDDAE0F39D CRC64;

Query Match 5.0%; Score 257.5; DB 1; Length 1162;
Best Local Similarity 20.3%; Pred. No. 0.0039;
Matches 233; Conservative 160; Mismatches 424; Indels 331; Gaps 52;

QY 34 YHNMSSSVVEKYLKLD-----INNLTQNLNLTYYKKSGRNKALK-- 73
DB 109 YENNTEDYQNTYLLSKNNEHYTANLVIFPGSGNIIKNVYIKKYEASGGMGTMLETW 168
QY 74 -----KFKELYTMVELEL-----KNNSLTPVEKNL 98
DB 169 FQPLFTHKYDEFVVDPALELIKLSLYLYGIKPNDNLNIPYRLRNEFNSLEYSELNW 228
QY 99 HTFIWGGQNDTAINYNOW-----KQVNSDYTVKVFVDSNAFLINTLUKTTI 145
DB 229 IDFLISGGIDYKLLNTNPYWFIDKPYDTSKNPEKYNDEYIKI--KNNNYIANSIKLYL 286

```

RESULT 6
MLP1_YEAST
ID MLP1_YEAST
AC Q02455;

STANDARD; PRT; 1875 AA.

[illegible]

Qy	426	TTMKIFSDKLIASINENDDMMFMKIKNYLVKVGAPADVRSTINISGPGVYTGAVQDLMFK	485
Ds	731	---KI-----IWNKDKMF-----PSFSKNSVL-----DTEFFN	756
Qy	486	DNSFTNHLPELRNFFPKTKISOLTEQBITSLWS-----ENQARAKSOFEEYKKGYPE	540
Ds	757	KKSNIITV-----IPDKLKKIQLNFQSIQKILNCFSLFFSKNKK-KTKIFQNSYF-	807
Qy	541	GALGEDNLDFAQNTVLVDKDYVSKKILSMKTRNKYIHYIVOLQSDGISYEASCNLF	600
Ds	808	-----INENLTTFSENDKEFNIFPLELFISEIINNDFLMRFFKK	846
Qy	601	DPYSSILYKQNIEGSETAYYYYADABI---KEIDKYRIPQISKNRNIKLTFTGHGKSEF	658
Ds	847	-----YLYRYIYKDKBEILFNPIENROLLQNFFETKTI-LTFI-----DF	884
Qy	659	NTDFFANLVDVDSLSSSETILNLAKADISPKYTEINLLGCNMFSSYSIYABETYPGKLLK	718
Ds	885	LOD-----PELNNRPFIFHEKTKIKN-----NLL-----YRLUKIP	919
Qy	719	IKDRVSELMPSISQDSITVSANOVEYRINBEGKREILDHSGKWINKEESIIKOISKEYI	778
Ds	920	LKDKRNFLL-----INE-----IK-----	933
Qy	779	SFNPKENKIIVKSK-----YLHELSTLLOETRNNANSDDIDLEK	817
Ds	934	SFIEKKNLFIKSQLSNVLVKNYSKFPFDNIFNFHLKQKEKNEIEILNNQNPESKLLK	993
Qy	818	KVMLTECEINVA-SNIDRQIVEGRIEAKNLSDSINYIKNEPKLIESISDSLY-----D	871
Ds	994	KTYLKNLNNNSYKPSYKIF---IPOLLNINL-----KNYKTFOWISELIFYSKNLN	1044
Qy	872	LKHQGLDDSHFTSFEDISKTENGFR-IRFINK-----ETGNSIFITEKEIFSEYATHIS	926
Ds	1045	YKIQNKIEKNYCNKNISYKKKKIKITVNFPEKKNLFTQNSWFFTLE---WWEYNTYIL	1101
Qy	927	KEISNIDKTIQDNVN-----GKLVKKNVLDAAHEVNTLNSAFFIOSLIEYNTTKESLS	979
Ds	1102	LQI---IQETFFQITDVLEYPKKKKIIEK-----NLKPFLLS-----KKISLK	1141
Qy	980	NLS	982
Ds	1142	TLS	1144
RESULT 8			
MSPI	PLAFK	STANDARD;	PRT; 1630 AA.
AC	P04932,		
DT	13-AUG-1987	(Rel. 05, Created)	
DT	01-FEB-1996	(Rel. 33, Last sequence update)	
DT	01-OCT-1996	(Rel. 34, Last annotation update)	
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens)		
DE	(PMWSA) (P190).		
GN	MSP-1.		
OS	Plasmodium falciparum (isolate K1 / Thailand)		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5839;		
FP	[1]		
RN	SEQUENCE FROM N.A.		
EX	MEDLINE=861136024;	PubMed=3004972;	
RA	Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,		
RA	Stunnenberg H., Bujard H.;		
RT	"Polymorphism of the precursor for the major surface antigens of		
RT	Plasmodium falciparum merozoites: studies at the genetic level.";		
RL	EMBO J. 4:3823-3829(1985).		
RN	[2]		
RP	REVISONS, SEQUENCE FROM N.A.		
RA	Pan W., Tolle R., Bujard H.;		
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.		
CC	-1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor		
CC	(Potential).		

CC -|- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X03371; CAA27070.1; -
CC InterPro: IPR006209; EGF_like.
CC Pfam: PF00008; EGF_1.
CC Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
CC Transmembrane; GPI-anchor.
CC FT SIGNAL 1 19 POTENTIAL.
CC FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
CC FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
CC FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
CC FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 955 955 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 1630 AA; 187289 MW; ADBDEC3C80A46322 CRC64;

Query Match 4.8%; Score 249; DB 1; Length 1630;
Best Local Similarity 20.9%; Pred. No. 0.012;
Matches 257; Conservative 215; Mismatches 418; Indels 338; Gaps 69;

QY 16 KPIQEDVEY-VAILN-ALEBYHNMSSVVEKYL-----KLKDINNLTNYLNTYK----- 64
DB 266 KKXLYQAQDLSYVQLQEAHNL--ISVLEKRIIDLTKKNIKELLDK-INEIKNPPPA 322
QY 65 KSG-----RNKAL-----KKFEY-----LTWEVLEL-----KN----- 88
DB 323 NSGNTPTNLDKNNKIEEHEKEIKETAKTIKFNIDSLFTDPLEYLYLREKKNIDISAK 382
QY 89 -----NSLT-PVEKHLHFTWIGGQINDTAINYNOWKDVNS--DYT---VKF 129
DB 383 VETKESTEPNEPNGVTYPLSYN-----DINN-ALNELNSFGDLINPFDYTKPESKN 433
QY 130 VYDSN-AFLINTLKTKT-----VPSATNNTLESFRENLDNPEFDYKPKRMETIYD 181
DB 434 IYTDNERKKFEINEIKIEKKIEKKIES-DKKSVEDRSKSLN-----DITKEYEKLNEIYD 488
QY 182 -KQKHFDY--YKQTEENPEFIDNII--KTYLSNEYSK-DIEALN---KYTEE-SINK 231
DB 489 SKFNNDIDLTFNFKMGKRYSVYKELTHHTTTFASYENSKHNLEKTKALKYMEDYSLRN 548
QY 232 ITANN-----GNDIRNLEKFADEDLRLYNQELVERWN-----LAAASDILRI 274
DB 549 IYVEKELKYKNLISKIENIEFELTVENIKDEQLPEKTKIDENKPEKILEVSDIVKV 608
QY 275 SM-----LKEDGGVLDVILPGIQ-PDLFKSINKPDSITNTSWEMIKLBAINK 322
DB 609 QVQKVLMMNKIDELKKTQLIKNVELKHNHVPNSYKQENKQBPY-----YLVLKKEIDK 664
QY 323 YKSYIPCYTSKFMDLDEE---VQSFESALSSEKSKSEI-----FLPLDD 365
DB 665 LKVFMPKVES-----LINEEKKNIKTEGQSDNSFSTEGEITGQATTKPQQAGSALEGDS 720
QY 366 IKVS-----PLEVKI-----AFANNSVINQALISLKSYSCLDVINQIKN 405

DB 721 VQAQAEQKQAPPPVPVPEAKAQVTPPPAPVNNKTENVSKLDYLEKLYEPLNTSYICH 780
QY 406 RYKILNDNLNPSINEGTFTNTMKIFSDKLASINEDNMFMKITYNLVKGFADVRST 465
DB 781 KYILVS---HSTMNE--KILQYKITKEESKSLSCDPLDLLFNITQNNIPVMYS--MFDS 833
QY 466 INLSGPGVYTGAYQ-----DILLMFKDSTNIHLPELNPFPKTKIS-----QLTEQ 514
DB 834 LNNLSUQLFMELIYEMVNCNLYKLDNDKIKNLLBEAKKVSSTVKTSSSSMQPLSLTPQ 893
QY 515 EITLSWNSFNQARAKSQFEYKKGYPEGALGEDNDLDFAQNTVLDK--DYVSKKILSSMKT 572
DB 894 DKPEV-SANDDTSHSTNLNLSLKFENILSLGKNKIYQELIGQSSSENFYEKILKDSST 952
QY 573 RNKEIYHIVOLQDKISYEASCNLFSDPYSSILYQKNIEGSETAYYYYVADAIKEID 632
DB 953 FYNESFTNFVKSADDIN---SLNDESK-----RKKLE-----EDIN 986
QY 633 KYRIPYOIS---NKRNIKLTFIGHKGSEFNTDTFANLDVDSLSSEIETILNLAKADISP 688
DB 987 KUKTKLQSLFDLYNKYKLERLFDKKTGVGKYKMKIKKLTLLKEQLESKLSLN---NP 1043
QY 689 KYIEINLLGCMFYSIY-----ABETYPG-KLLAK-----IKDRVSELMPP--S 729
DB 1044 KHV-----LQNFVFPNKKKEAEIAETENTLENTKILKHVKGLVKKYNGESSPLKT 1095
QY 730 ISQDSITVAN-----QYEVRIINEGK-----REILDHSGKWINKESIIKD-ISK 775
DB 1096 LSEESIOTEDNYASLENFKVLKLEGKLDNLEKKLSYLSGLHHLIAELKEVIRNK 1155
QY 776 EYISNPKENKIIVKS-----KYLHE---LSTLQEIERNANSDDILEKKVMLTEC-- 824
DB 1156 NYTGSPSENNTDVNNALESYKFLPEGTDVATVYSE--SGSDTLLEQSQPKPASTHVA 1213
QY 825 ---EINVASNIDRQI-----VEGRIBEAKNLTSDSINYIKNEFKLIESIDSPLYDLKH 874
DB 1214 ESNTITTSQNVDDVDVLIIPFGESE---DYDDLQGVVTGEAVTPSVIDNI----- 1264
QY 875 QNGLDDSHFISPEDISKTEGFRIRINKEGNSIFETERKEIFSEYATHISKEISNIKD 934
DB 1265 -----LSKIENEYEVLYKPLAG--VYRSLLKQLENNVMTF-----NVNVK 1304
QY 935 TI-----PDNV-NGKLVKKNLDAAEV-----NTLASAFFIQLIJEYNTTKESL- 978
DB 1305 IUNSFNKNRENKPNVLESOLIPYKDLTSSNYVVKDPYKFLNKEKDKFLUSSNYIKDSID 1364
QY 979 SNLSVAMKVQVYAQLFSTGLNTITDASK 1006
DB 1365 TDINFANDVLGYKYLSEKYSDDLDSIK 1392

RESULT 9
MSPL_PLAFW
ID MSPL_PLAFW STANDARD; PRT; 1639 AA.
AC P04933;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMWSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5848;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86014355; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA Nicholas S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.F.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites.";

RL Nature 317:270-273 (1985).
RN (2)
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/>
CC OR SEND AN EMAIL TO license@isb-sib.ch).
CC
CC EMBL; X02919; CAA26676.1; -.
DR PIR; A24594; A24594.
DR PIR; S05603; S05603.
DR PDB; 1CEJ; 28-MAY-99.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor; 3D-structure.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 764 764 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 768 768 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;

Query Match 4.8%; Score 249; DB 1; Length 1639;
Best Local Similarity 20.9%; Pred. No. 0.013;
Matches 257; Conservative 215; Mismatches 418; Indels 338; Gaps 69;

QY 16 KFRIOEDY-VAILN-ALBEYHNMSESSVVEKYL-----KLKDINNLTNLYNTYK----- 64
DB 275 KKKLYQAQYDLSYNNKQLEAHNL--ISVLEKRIDTLKKNENIKELLDK-INEIKNPPPA 331
QY 65 KSG-----RNKAL-----KFPKEY-----LTMEVLEL-----KN----- 88
DB 332 NSGNTPTLLDKNKKIEEHEKEIKETAKTIKFNIDSLFTDPLEYLEYLRKKNKNDISAK 391
QY 89 -----NSLT-PVEKNLHPIWGGOINDTAINYINQWQVNS--DYT----VKF 129
DB 392 VETKSTENPVGVTYPLSYN-----DINN-ALNELSFGDLINPFDTYKPSKN 442
QY 130 VYDSN--AFLINTLKKTI-----VESATNNTLESRENLDPEFDYKFKRMEIYD 181
DB 443 IYTDNERKKFINEIKIEKIEKKIES-DKSYEDRSKSLN-----DITKEYKLLNEIYD 497
QY 182 -KQKHPIDY--YKSOITEPEFIIDNII--KTYLSNEYSK-DLEALN-----KYIEE-SLNK 231
DB 498 SKFNNNIDLTNPEKMGKRYKYVEKLTHHTTFASYENSKHLEKLTALKATKYMEDYSLRN 557
QY 232 ITANN-----GNDIRNLEKFADEDLVRLYNQELVERWN-----LAAASDILRI 274
DB 558 IYVEKELKYKNKLNISKIENIEITLVENIKKBEQLEFKEKTKIDENKDPDKILEVSDIVK 617
QY 275 SM-----LKEDGGVLDVDILPGIQ-PDLFKSINKPDSITNTSWEMIKLEAIMK 322

DB 618 QVQKVLANKIDELKKTQLILKNVELKNIHVPNSYKQENKQEPY-----YLVLKKEIDK 673
QY 323 YKEYIPGYTSKNFMDLDEE---VQSPESALSSKSDKSEI-----FLPLDD 365
DB 674 LKVPMPKVES-----LINBEKKNITGEGSDNSEPSTEGEITQCATTKPGQQAAGSLEGS 729
QY 366 IKVS-----PLEVKI-----AFANNVINQALISLKDYSYCSDLVINQIKN 405
DB 730 VQAQAQEQQAQPPVPVPPEAKAQVPTPPAPVNNKTNVNSKLDYLEKLYEFLNTSYICH 789
QY 406 KYKILNDNLNPSINRGTDFTMTKIPSDKLASINEDNMFMKITYNLYKGFADPVRST 465
DB 790 KYILVS---HSTMNE--KILQYKITKEESKSLSCDPLDLNFIQNNIPVMYS--MFDS 842
QY 466 INLSGPGVVTGAYO-----DILLMPKDNSTNHLLEPELRNPFEPKTKIS-----QLTSQ 514
DB 843 LNNLSQJFMEIYEKVMCVNLYKLDKNDKIRNLEEKVKVSVTKLSSSSQPLSLTPQ 902
QY 515 EITSLWSFNQARAKSQPBKYGYPGALGEDDNDLDFAQNTVLDK--DYVSKILSSMKT 572
DB 903 DKPEV-SANDDTSHSTNLNLSLKLFENILSLGKNKIYQELIGKSSSENFYKILKDSOT 961
QY 573 RNKEVIHVIVOLGQDKIYBASCNLFSDPYSSILYQKNIEGSETAYYYYVADAIEKID 632
DB 962 FYNESFTNFVKSADDIN---SLNDESK-----RKKLE-----EDIN 995
QY 633 KYRIPYQIS---NKENIKLTFIGHKSEFNFTDTPANLDVDSLSSEIETILNLAKADISP 688
DB 996 KLKTLQLSFDLYNKYKLERLFDKKTGVGYKMQIKKLTLLKQLESKLSLN---NP 1052
QY 689 KYIEINLLGCNMFYSIY-----AETYPG-KLLK-----IKDRVSELMW--S 729
DB 1053 KHV-----LQNESVFNKKKEAEIAETENTLENTKILLKHYKGLVKNYNGESSPLKT 1104
QY 730 ISQDSITVSAN-----QYEVRIINEGK-----REILDHSGKWINKESIIKD-ISK 775
DB 1105 LSEESIQTEDNVASLENFVKLSKLGKLDNLEKKLKSYLSSGLHLHLIAELKEVIKKN 1164
QY 776 EYISFNPKENKIIVKS-----KYLHE---LSTLQEIERNANSSDIDLEKKVMLETC--- 824
DB 1165 NYTGNPSENNNTDNNALLESYKFLUPEGTDAIVATVSE--SGSDTILEQSQPKPASTHVA 1222
QY 825 ---EINVASNIDROI-----VEGRIBEAKNLTSDSINYIKNEFKLIESIDSIDYLDKH 874
DB 1223 ESNTITTSQNVDDVDVIVIPFGESEB---DYDDLQGVVTGEAVTPSVIDNI----- 1273
QY 875 QNGLDDSHFISFEDISKTEGFRIRFINKETGNSIFIEKEIFSEYATHISKEISNIKD 934
DB 1274 -----LSKIENEYEVLYLKPLAG--VYRSLLKKQLENNVMTF-----NVNVKD 1313
QY 935 TI-----PDNV-NGKLVKNVLDAAHEV-----NTLNSAFFIQSLIEVNTTKESI- 978
DB 1314 ILNSRFNKNFKNVLESDLIPYKOLTSNYYVVDPPYFLNKEKRDKFLSSNYIKDSID 1373
QY 979 SNLSVAMKVQVYAQLFSTGLNTITDASK 1006
DB 1374 TDINFANDVLGYKILSEKYSGLDSIK 1401

RESULT 10
Y109 YEAST
ID Y109 YEAST STANDARD; PRT; 1679 AA.
AC P40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 195.1 kDa protein in DNA43-UBI1 intergenic region.
GN Y1149C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC
CC
DR EMBL: M19143; AAA29653.1; --
DR PIR; A54498; A54498.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193719 MW; 3920875E73D38552 CRC64;

Query Match 4.7%; Score 245; DB 1; Length 1701;
Best Local Similarity 20.1%; Pred. No. 0.019;
Matches 213; Conservative 176; Mismatches 370; Indels 300; Gaps 51;

QY 36 NMSESVVVEKYLKLDINNLDTNLYTKSGRNKALKKFKFVLTWVLELKNLSLTPVE 95
DB 106 NFSNDSNSTTKYADLKHVQNYLFT-----IKELKYPFLDLTNHMLT-LS 152

QY 96 KNLFIWICGQNDRAINTVINOQKOVNS--DYTVKFDVDSNAFLNTLTKTIVESATNWL 154
DB 153 KNV-----DGFYLDGVEEINELLYKLNFFYD-----LLRAKLDACANSYC 195

QY 155 E-SFRENLDPPFDVYK-----FVRKMEIYDKQHFIDYKQSQIEENPEFIIDNIKTY 209
DB 196 QIPFNLIKIPANELDVLKIVFGYRKLPLDNKKNVGMKEDYIKKN-----KTT 242

QY 210 LSNEYSKOLEALNKYIEESLNKITTANNNGNDIRNLEK--FADEDLVRLYNQELVERNLAA 267
DB 243 IAN-----INELIEGSKTIDQNKADNEEGKKLYQAVNYLQKQEAHNLIS 294

QY 268 ASDILRISLMKEDGGVLDVILPGIQLPKSI--NKPDSTNTSWEMKLEAMKYKE 325
DB 295 VLE-KRIDTLKKNENIKKLEIDIKITDAENPTTGSKENPLPENK-----KK 341

QY 326 YIPGYTSK-----NFDMLDREVQVSFESALSSKSKSEIPLPLDDIKVSPLEVKI 375
DB 342 EVEGHEEIKETAKTKFNIDSLFTD-PLEEYLYREKKNKVDVTPKSDPTKSVQIPKV 400

QY 376 AFANNSVINQALISLSDYSCLVINGQIKNRY-KILNDNLNPSINSGTDFNTMTKIFSDK 434
DB 401 PYPNGIVYPLPLTDIHNLSAAD-----NDKNSYGLMNPDPTEKINE-----KIITD- 447

QY 435 LASISNEDNMFMKITYNLKVGFPADVRSTNLSGPGVYTGAYQDLLMFKONSTNIHLL 494
DB 448 -----NKRKIFI-----NNIKKQIDL-----464

QY 495 EPELRNFEFPKTKISLTQETISLWSFNOARAKSQFEYKKGYPREGALGEDNDLDAFON 554
DB 465 -----EERKNINTKQNKKLE-----DYKSKDYBELLEKFPYB-----MKPNNN 505

QY 555 TVLDDKYVSKKTLSSMKTRNKE---YIH-----YIVLQGDGKISYEASCNLFSKDP 602
DB 506 --FDKDVVD-KIFSARYTYNVEKQRYNNKFSSNNNSVYVQKALKALSY-----LSD 554

QY 603 YSSILYQKNIEGSETAYYYYV---ADAETIKDKVIRPQISNKRN--IKLTFIG--HGK 655
DB 555 YS---LRKGISEKDFNHYYTLTKGLEADIKKLE-----EIKSSENKILEKFNKGLTHSA 606

QY 656 SFENTDTFANLDV-DSLSEIETILNLAADISPKYIEINL-----LGCNMFPSYSIYAET 710
DB 607 N-----ASLEVSIVLQVQKVLIIKKIEDLRK-IELFLQNAQLKOSIHVPNIYKPQN 658

QY 711 YPGK-----LLLKIKDRVSELMPISIQ-----DSIT---VSANQYEVVRINEBG--KR 752
DB 659 KPPEYLVILVKKEVDKLEKFIKPKVDMKLEQAVLSSITQPLVAASE-----TTDCGHST 714

QY 753 EILDHSGKWINKEESIIKD-----ISSKEYISFNKPKENKIIVKSKYLHELSTL 800
DB 715 HTLSQSGETEVEETEVEETVGTHTTTVTITLPPKEBSA--PKEVKVWNS-----763

QY 801 LQEIENNANSSDIDLEKKVMLTECEINVASN-----IDRQIVEGRIEERAKNLTSDSI 852
DB 764 ---IEHKSNDNSQALTKTVYKCLDEFLTKSYCHYILVSNSSMDOKLLEVNLTPBEE 820

QY 853 NVYKNEFKL-----IESISDSLYDLKHQGLDDSHFISPFEDISKTENGPRIRINKETON 907
DB 821 NELKSCDPLDLLFNITQNNIPAMYSLYDSNMNIDLOHL--PFELYOKEMIYLLHKLKEENHI 878

QY 908 SIFIFETEKEI-----FSEYATHISKEISNIKDTIFDNVNGKLVKK--VNL 950
DB 879 KLLBEQKQITGTSSTSPGNTVNTAQSATHSNQOOSNASTNTQNGVAVSSGPAVV 938

QY 951 DAAHEVNTLNSAFFIQSLIEYNTTKESLSNLSVAMKVQV 989
DB 939 EBSHDFLIVLSI-----SNDLKGIVSLNLGNKTKV 969

RESULT 12
MSPI_PLAF3
ID MSPI_PLAF3 STANDARD; PRT; 1682 AA.
AC P19598; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMWSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5834;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE=8816657; PubMed=3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen precursor p190 of Plasmodium falciparum lacks tripeptide repeats."; EMBO J. 6:4137-4142(1987).
RN [2]
RP SEQUENCE OF 1032-1682 FROM N.A.
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1."; Exp. Parasitol. 81:47-54(1995).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; 235326; CAA84555.1; --
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merzoite; Polypotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL .1 19
FT CHAIN 20 1682
FT TRANSMEM 1666 1682
FT CARBOHYD 233 233
FT CARBOHYD 462 462
FT CARBOHYD 528 528
FT CARBOHYD 599 599
FT CARBOHYD 785 785
FT CARBOHYD 881 881
FT CARBOHYD 901 901
FT CARBOHYD 947 947
FT CARBOHYD 1071 1071
FT CARBOHYD 1178 1178
FT CARBOHYD 1569 1569
SQ SEQUENCE 1682 AA; C82A1E159948CAD6 CRC64;

Query Match 4.7%; Score 244; DB 1; Length 1682;
Best Local Similarity 19.8%; Pred. No. 0.02;
Matches 207; Conservative 180; Mismatches 374; Indels 284; Gaps 49;

QY 36 NMSESSVVEKYLKLDINLTNYLTKYSGRNKALKKFKVLTMEVLELKNNSLTPVE 95
DB 100 NPSDDSSDAKADKRVQNYLFT-----IKELKYPELFDLTHMLTCLD 147

QY 96 KNLHFIWIGQINDTAINY-INQKDVNSDYTVKPYVDSNAFLINTLKKTIIVESATNNTL 154
DB 148 -NIH-----GFKYLDIGYBEIN-----ELLYKLN-FYDOLLRAKLDVNCANDYC 189

QY 155 E-SFRENLDPEPDYK----FYKRMELIYKQKHFIIDYKQSBENPEFIIDNIKY 209
DB 190 QIPFNLKIRANELDVLKLVFGYRKPLDFIKNVGRMEDYIKKN-----KTT 236

QY 210 LSNEYSKDLALNKYTEESLNTKNTANGNDIRNLEK---FADEDLRLVNOELVERWNL 266
DB 237 IAN-----INLEGGKKTIDQKNADNBEGKKLVQAOVDLF-IYNNKQLOEHNLI 287

QY 267 AASDILRLSMLEKGGVYLDVILPGIQDLPKFSINKPDSITNTSMWKLEAIKMYKEY 326
DB 288 SVLE-KRIDTLKKNENIKLLEDIDKIDAEKPTTGVNQILS-----LRLEKESRHEBK 341

QY 327 IP--GWTSK-NFMDLDEEVQSPESALSSKSEIFLPLDDIKVSPLEVKAFAPANNVI 383
DB 342 IKELIAKTIRFNIDRLFTD-PLELEYLREKNKKVDVTPKSDQPTKSVQIPKVPYNGIY 400

QY 384 NQALISLKDYSYCSLDVINGIKRY-KILNDNLNPSINEGTFDNTTWKIFSDKLASINED 442
DB 401 PLPLTDIHSLSAAD-----NDKNSYGLMMPHTYKEKINE-----KIITD-----NKE 442

QY 443 NMFMKIKITNYLKVGPAPDVRSTINLSGPGVYTGAYQDILLMFKDSTNIHLLEPELRNPE 502
DB 443 RKIFI-----NNIKKQIDL-----EKNIN 462

QY 503 FPKTKISQTEQBITSLWFSNQARAKSQFEYKKGFGEGALGEDDNLDPQNTVLDKOVY 562
DB 463 HTKEQNKGLLE-----DYKSKQYELLEKEFYEMKFNFNK-----VVDKIFS 508

QY 563 SKKILSMKTR-NKEYIH-----YIVQLOGDKISYEASCNLSFKDPSYILYQKIEGSE 616
DB 509 ARTYNVEKQRYNNKFSNNSSNVNVQKLKALSY-----LEDYS---LRKGISEKD 557

QY 617 TAYYYV---ADAEIKEDIKRIPYQISKNRN---IKLTFIG--HGKSEFTDTFANLDV- 668
DB 558 FNNHYTLTKGLEADIKKLETE-----EIKSENKILEKNFKGLTHSAN-----ASLEVS 605

QY 669 DLSSEIEITLNAKADISPKYIEINL-----LGCNWFYSYIYAETYPCK-----LLLKIK 720
DB 606 DIVKLOVQVLLFKKIEDLRK-IEFLKNAQLKDSIHVPNIYKPNQKPEFYLIIVLKKEV 664

QY 721 DRVSELMF-----SISQDITSVANQYE-----VRINEGKRIILDHSGKW 761
DB 665 DKLEFIPKVKDMLKKEQAVLSSITQPLVAASETTEDGGHSTHTLSQSGETEVTETEET 724

QY 762 INKEESIIOKDISSEKEYISFNPKENKIIVSKYKYLHSLSTLLOEIRNANSSDIDLEKKVYL 821
DB 725 VGHITTVT-----ITLPPKEVKVENS-----IEHKSNDNSQALTKTIVYL 764

QY 822 TCEINVASN-----IDRQIVEGRIEEAKNLT-----DSINYIKNEFKLI 862
DB 765 KKLDFLTKSYCHYKYLIVSNSSMDQKLELVNLTPEENELKSCDRDLDFLFIQNNIPAM 824

QY 863 ESISDSL-YDLKHONGLDSDHSPISFEDISKTEGFRIFRINKETGNSIFIEKEI---- 917
DB 825 YSLYSMDNDLQ-----LFELYQEMYYLHLKEENHIKKLEEPKQITGTS 874

QY 918 -----FSEYATHISKEISNIKTIFDNVNGKLVK--VNLDAAHEVNTLNSAFF 964
DB 875 STSSPGNTTNTAQSATHSNQSQSNASTNTQNGVAVSSGPAVVEESHDLPLVLSI-- 932

QY 965 IQSLIEYNTTKESLSNLSVAMKQV 989
DB 933 -----SNDLKGIVSLNLGNKTKV 951

RESULT 13
MSPL_PLAFM
ID MSP1_PLAFM STANDARD; PRT; 1701 AA.
AC P08569;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Merzoite surface protein 1 precursor (Merzoite surface antigens)
DE (PMSEA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=70153;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite
Plasmodium falciparum".
RL J. Mol. Biol. 195:273-287(1987).
RP [2]
REVIEWS TO 1403; 1569 AND 1629.
RA Tanabe K.;
RN Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RX [3]
SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
Plasmodium falciparum merzoites: studies at the genetic level".
RL EMBO J. 4:3823-3829(1985).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
(Potential).
CC -!- PTM: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X05624; CAA29112.1; --
DR InterPro; IPR006209; EGF_like.

DR	Pfam; PF00008; EGF, 1.	
KW	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;	
KW	Transmembrane; GPI-anchor.	
FT	SIGNAL	1 19
FT	CHAIN	20 1701
FT	CARBOHYD	110 110
FT	CARBOHYD	239 239
FT	CARBOHYD	470 470
FT	CARBOHYD	536 536
FT	CARBOHYD	607 607
FT	CARBOHYD	802 802
FT	CARBOHYD	899 899
FT	CARBOHYD	919 919
FT	CARBOHYD	965 965
FT	CARBOHYD	991 991
FT	CARBOHYD	1089 1089
FT	CARBOHYD	1196 1196
FT	CARBOHYD	1588 1588
FT	SEQUENCE	1701 AA; 193768 MW; 3FC2EC59AP96EA98 CRC64;
Query Match	4.7%; Score 244; DB 1; Length 1701;	
Best Local Similarity	20.6%; Pred No. 0.02;	
Matches	220; Conservative 174; Mismatches 357; Indels 316; Gaps 54;	
QY	36 NMSESVVEKYLKLDINNLTNYLNTYKSGRNKALKKFKYLTWVLELKNLSLTPVE	95
DB	106 NPSNDSNSTTKYADLKHVQNYLFT-----IKELYPFLDLNHLT-LS	152
QY	96 KNLHFIWGGQINDTAINTVINOQVNS-DYTVKVFYDSNAFLNTLTKTIVESATNTL	154
DB	153 KNV-----DGFYKYLIDYSEINELLYKLFYD-----LLRAKLNDACANSYC	195
QY	155 E-SRENLDNPPDFYK-----FYRKMEIYIKQKHFDYKYSQIEENPEFIIDNIKT	209
DB	196 QIPFLNKRANLDELVKLVFGYRPLDNKKNVGMEDYIKN-----KIT	242
QY	210 LSNEYSKDLALNKYIEESLNKITANGNDIRNLEK--FADEBLVRLYNQELVERNLAA	267
DB	243 IAN-----INELIEGSKTIDQNKADNEEGKKLYQAOVNLFIYNKQLQEAHLIS	294
QY	268 ASDILRISMLKEDGGVYLDVILPGIOPPLFKSI--NKPDSTNTSWEMIKLEAMTKYKE	325
DB	295 VLE-KRIDTLKKNENIKKLEIDIKTDAENPTTGSKENPLFNK-----KK	341
QY	326 YIPGYTSK-----NFDMLDEVOVSFSSALSSKSDKSEIFLPDLDIKVSPLEVKI	375
DB	342 EYEGHEEKIKETAKTIKFNIDSLFTD-PLELYYLRKNKRVDTVPKSDPTKSVQIPKV	400
QY	376 AFANNSVINQALISLKDSCSLVINQIKNRY-KILNDNLNPSINEGTDFTNTMIFSKD	434
DB	401 PYPNGIVYPLPLTDIHNSLAAD-----NDKNSYGDLMNPDTKEKINE-----KIITD-	447
QY	435 LASIGNEDMMFMKIKITNLYKVGPDVVRSTINLGGVYVYGAYQDLLMFKDNSTNIHLL	494
DB	448 -----NKRKIFI-----NNIKQIDL-----	464
QY	495 EPELNFPEPKTKISOLTQBITSLWSFNQARAKSQFEYKGYEGALGEDDNLDAQON	554
DB	465 -----EERNINHTKQNKILLE-----DYKSKDYELLEKFYE-----MKFNN	505
QY	555 TVLVDKDYVSKLSSMKTANKE--YIH-----YIVLQGGOKISYEASCNLFPSKDP	602
DB	506 --FDKDVVD-KIFSARYTYNVEKQRYNNKFFSSNNNSVNVQKLKALS-----LED	554
QY	603 YSSILYQKNIEGETAYIVYV--ADAETIKEDKVRIPYQISKNR--IKLTFIG--HGK	655
DB	555 YS--LRKGISKDFNHYYTLTKGLEADIKKLT-----EIKSSENKILEKKNFKGLTHSA	606
QY	656 SEFNTDTFANLDV-DSLSSEIETILNLAADISPKYIEINL---LGCNMFYSYSYABET	710
DB	607 N-----ASLEVSQDILVQLQVOKVLIKIEDLRK-IEFLFKNAQLKDSIHVPNIYKPN	658
QY	711 YPGK-----LLLKIKDVRSELMPISQ-----DSIT---VSAHQYEVRIINEEG--KR	752

DB	659 KPPEYLLIVLKKEVDKLEKFIKPKVKMDLKEQAVLSSITQPLVAASE-----TTEDGGHST	714
QY	753 EILDHSGKWINKESIIND-----ISSKEYISFNPKNENKIIVKSKYLHELSTL	800
DB	715 HTLSQSGTEVTEVTEVTEVTEVTHVTITVITLPPKBSA--PKVKVNVNS-----	763
QY	801 LQERNNANSDDIDLEKKVMLTECE-----INVA-SNIDRQIVEG---RIBEA	844
DB	764 ---IEHKSNDNSQALTKTVYKLDLDEPLTKSYCHKYLVSNSNDQKLLVYVNLTPEE	820
QY	845 KNLTS-----DSINYIKNEFKLIESISDSL-YDLKHQNGLDDSHFISPDISKTENGFR	899
DB	821 KELKSCDPLDLLFNQNNIPAMYSLYDSMNNDLQK-----LFFELYQKEMIIYVLH	870
QY	900 FINKETGNSIFETEKET-----FSEYATHLSKEISNIKOTIFDNVNGKL	944
DB	871 KLKEENHIKKLEBQKQITGTSSSTSPGNTVNTAQASATHNSQNSQNASSTNTQNGVA	930
QY	945 VKK--VNLDAAHEVNTLNSAFFIQSLIEYNTTKESLNSLSVAMKVQV	989
DB	931 VSSGPAVVEESHDPILTVLSI-----SNDLKGVLSLLNLGNKTKV	969
RESULT 14		
ID	MSPI_PLAFC	
AC	P04934; STANDARD; PRT; 1726 AA.	
DT	13-AUG-1987 (Rel. 05, Created)	
DT	01-MAR-1989 (Rel. 10, Last sequence update)	
DT	01-OCT-1996 (Rel. 34, Last annotation update)	
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens)	
DE	(PMMSA) (P195).	
GN	MSP-1.	
OS	Plasmodium falciparum (isolate Camp / Malaysia).	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI TaxID=5835;	
RN	[1]	
RP	SEQUENCE OF 1-1103 FROM N.A.	
EX	MEDLINE=86205236; PubMed=3517809;	
RA	Weber J.L., Leininger W.M., Lyon J.A.;	
RT	"Variation in the gene encoding a major merozoite surface antigen of	
RT	the human malaria parasite Plasmodium falciparum."	
RL	Nucleic Acids Res. 14:3311-3323(1986).	
RN	[2]	
RP	SEQUENCE OF 1104-1726 FROM N.A.	
EX	MEDLINE=86143999; PubMed=3378296;	
RA	Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;	
RT	"Merozoite surface protein sequence from the Camp strain of the human	
RT	malaria parasite Plasmodium falciparum."	
RL	Nucleic Acids Res. 16:1206-1206(1988).	
CC	-!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor	
CC	(Potential).	
CC	-!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42	
CC	kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF	
CC	MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
CC	EMBL; X03831; CAA27446.1; -.	
DR	PIR; A23386; SAZQCM.	
DR	InterPro; IPR006209; EGF_like.	
DR	Pfam; PF00008; EGF; 1.	
KW	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;	
KW	Transmembrane; GPI-anchor.	
FT	SIGNAL	1 19
FT	CHAIN	20 1726
FT	SEQUENCE	1726

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:26:27 ; Search time 30 Seconds
(without alignments)
8773.789 Million cell updates/sec

Title: US-09-126-816b-6_COPY_1_1020

Perfect score: 5192

Sequence: 1 MNLVNAQLQKVMYVKFRITQ.....ITDASKVVELNSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5171	99.6	2364	2 Q46342	Q46342 clostridium
2	4090.5	78.8	2367	2 Q9EXR0	Q9EXR0 clostridium
3	4081.5	78.6	2367	2 Q9F931	Q9F931 clostridium
4	4073.5	78.5	2367	2 Q46034	Q46034 clostridium
5	2144.5	41.3	554	2 Q93139	Q93139 clostridium
6	1911	36.8	698	2 Q68653	Q68653 clostridium
7	1893	36.5	698	2 Q9EXQ8	Q9EXQ8 clostridium
8	1887	36.3	697	2 Q86141	Q86141 clostridium
9	1469	28.3	553	2 Q9FCX5	Q9FCX5 clostridium
10	1424.5	27.4	2178	2 Q46149	Q46149 clostridium
11	411.5	7.9	3169	2 Q9ZGR4	Q9ZGR4 escherichia
12	408.5	7.9	3169	2 Q82916	Q82916 escherichia
13	386	7.4	3223	2 Q8L1L9	Q8L1L9 escherichia
14	384	7.4	3223	2 Q9RPH1	Q9RPH1 escherichia
15	381	7.3	3223	2 Q9RM48	Q9RM48 escherichia
16	381	7.3	3223	2 Q8VQR2	Q8VQR2 escherichia

17	377	7.3	3223	2 Q8VNR6	Q8VNR6 escherichia
18	362	7.0	3225	16 Q9PKM6	Q9PKM6 chlamydia m
19	357.5	6.9	2771	5 Q26216	Q26216 plasmodium
20	352	6.8	3317	16 Q8EWP8	Q8EWP8 mycoplasma
21	346.5	6.7	2752	5 Q8BJY0	Q8BJY0 plasmodium
22	345.5	6.7	1127	12 Q9YV76	Q9YV76 melanoplus
23	345	6.6	5767	5 Q81525	Q81525 plasmodium
24	338	6.5	2204	5 Q81L70	Q81L70 plasmodium
25	337.5	6.5	2747	5 Q9BJX9	Q9BJX9 plasmodium
26	335.5	6.5	2867	5 Q9N2M3	Q9N2M3 plasmodium
27	330	6.4	3724	5 Q77320	Q77320 plasmodium
28	324.5	6.2	1238	12 Q9EMP3	Q9EMP3 amacta moo
29	318.5	6.1	2166	16 Q51465	Q51465 borrelia bu
30	316.5	6.1	3254	5 Q9BK45	Q9BK45 plasmodium
31	316.5	6.1	3255	16 Q9PKM8	Q9PKM8 chlamydia m
32	315.5	6.1	3130	5 Q81DX6	Q81DX6 plasmodium
33	315.5	6.1	3394	5 Q77384	Q77384 plasmodium
34	315	6.1	1453	8 Q9G9H3	Q9G9H3 schizophyll
35	315	6.1	3130	5 Q9BK46	Q9BK46 plasmodium
36	314.5	6.1	1387	5 Q9GZ76	Q9GZ76 plasmodium
37	313	6.0	3381	5 Q812V4	Q812V4 plasmodium
38	312	6.0	1956	5 Q81IE1	Q81IE1 plasmodium
39	312	6.0	5779	5 Q8IBS0	Q8IBS0 plasmodium
40	311	6.0	1302	2 Q49547	Q49547 mycoplasma
41	310.5	6.0	5687	5 Q81JH4	Q81JH4 plasmodium
42	310	6.0	2269	5 Q26223	Q26223 plasmodium
43	310	6.0	3335	16 Q9PKM7	Q9PKM7 chlamydia m
44	308	5.9	2792	5 Q814R2	Q814R2 plasmodium
45	305	5.9	1686	5 Q81FP9	Q81FP9 plasmodium

ALIGNMENTS

RESULT 1

Q46342	ID	Q46342	PRELIMINARY;	PRT; 2364 AA.
AC	Q46342;			
DT	01-NOV-1996	(TremBLrel. 01, Created)		
DT	01-NOV-1996	(TremBLrel. 01, Last sequence update)		
DT	01-MAR-2003	(TremBLrel. 23, Last annotation update)		
DE	Cytotoxin L.			
OS	Clostridium sordellii.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1505;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=6018;			
RX	MEDLINE=95369733; PubMed=7642137;			
RA	Green G.A., Schue V., Montell H.;			
RT	"Cloning and characterization of the cytotoxin L-encoding gene of			
RT	Clostridium sordellii: homology with Clostridium difficile cytotoxin			
RT	B.;"			
RL	Gene 161:57-61(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=6018;			
RX	MEDLINE=96149194; PubMed=8544213;			
RA	Green G.A., Schue V., Girardot R., Montell H.;			
RT	"Characterisation of an enterotoxin-negative, cytotoxin-positive			
RT	strain of Clostridium sordellii.;"			
RL	J. Med. Microbiol. 44:60-64(1996).			
DR	EMBL; X82638; CNA57959.1; -			
DR	InterPro; IPR002479; CW_binding.			
DR	InterPro; IPR001950; TIF_SUII.			
DR	Pfam; PF01473; CW_binding.1; 18.			
DR	PROSITE; PS01118; SUI1.1; 1.			
SQ	SEQUENCE 2364 AA; 270576 MW; EAD8A4467A89BDBB CRC64;			

Query Match 99.6%; Score 5171; DB 2; Length 2364;
Best Local Similarity 99.7%; Pred. No. 4e-195;
Matches 1017; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MNLVNAQLOKQVYVFRIOEDYVAILNALBEYHNMSSESVVEKYKLKLDINNTDNYL 60
DB 1 MNLVNAQLOKQVYVFRIOEDYVAILNALBEYHNMSSESVVEKYKLKLDINNTDNYL 60
QY 61 NTYKSGRNKALKKFKXYLTMEVLEKNSLTPVEKNLHFHWIGQINDTAINYNQWKD 120
DB 61 NTYKSGRNKALKKFKXYLTMEVLEKNSLTPVEKNLHFHWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKPYDSNAFLINTLTKTIVESATNTNLTESFRENLDNDFDYNKFKYKMEIY 180
DB 121 VNSDYTVKPYDSNAFLINTLTKTIVESATNTNLTESFRENLDNDFDYNKFKYKMEIY 180
QY 181 DKQKHFDYKSOIENPFIIDNIITKYLSEYKDLALNKYIEESLNKLTANNNDI 240
DB 181 DKQKHFDYKSOIENPFIIDNIITKYLSEYKDLALNKYIEESLNKLTANNNDI 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300
DB 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300
QY 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVORFESALSSKSKSEIF 360
DB 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVORFESALSSKSKSEIF 360
QY 361 LPLDDIKVSPLEVKIAFANNSVINOALISLKDSYCSDLVINQIKNRYKILNDNLNPSINE 420
DB 361 LPLDDIKVSPLEVKIAFANNSVINOALISLKDSYCSDLVINQIKNRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYQD 480
DB 421 GTDFNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYQD 480
QY 481 LLMFKDNSTNIHLEPELNFPPTKIISQTEQITSLWSFNQARAKSQFBEYKKGYPE 540
DB 481 LLMFKDNSTNIHLEPELNFPPTKIISQTEQITSLWSFNQARAKSQFBEYKKGYPE 540
QY 541 GALGEDNDLPQNTVLDKDYVSKLSSMKTRNKEYIHYIYVLOQDKLSYEASCNLPSK 600
DB 541 GALGEDNDLPQNTVLDKDYVSKLSSMKTRNKEYIHYIYVLOQDKLSYEASCNLPSK 600
QY 601 DYPSSILYQKNIEGSETAYYYVADAEIKEDKIRIPYQISNKRNIKLTFIGHGKSEFNT 660
DB 601 DYPSSILYQKNIEGSETAYYYVADAEIKEDKIRIPYQISNKRNIKLTFIGHGKSEFNT 660
QY 661 DTFANLDVDSLSSEIETILNLAADISPKYIEINLGCNMFYSYVAEETYPGKLLKTK 720
DB 661 DTFANLDVDSLSSEIETILNLAADISPKYIEINLGCNMFYSYVAEETYPGKLLKTK 720
QY 721 DRVSELMPISQDSITVSANOYEVRIENEGKREILDHSGKWINKEESIHKDISKEYISF 780
DB 721 DRVSELMPISQDSITVSANOYEVRIENEGKREILDHSGKWINKEESIHKDISKEYISF 780
QY 781 NPKENKIIVKSXYLHELSTLQEIERNANSDDILEKKVMTLCEINVASNIDRQIVEGR 840
DB 781 NPKENKIIVKSXYLHELSTLQEIERNANSDDILEKKVMTLCEINVASNIDRQIVEGR 840
QY 841 IEBAKNLTSDSINYIKNEFKLIESISDSLYDLKHQNGLDDSHFISPEDISKTENGFRIRF 900
DB 841 IEBAKNLTSDSINYIKNEFKLIESISDSLYDLKHQNGLDDSHFISPEDISKTENGFRIRF 900
QY 901 INKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHVNTLN 960
DB 901 INKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHVNTLN 960
QY 961 SAFFIQSLYEYNTKESLSNLSVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 SAFFIQSLYEYNTKESLSNLSVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDETID 1020
```

RESULT 2
Q9EXR0

PRELIMINARY; PRT; 2367 AA.

QY

```
Q9EXR0;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
Toxin B.
DE TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA von Eichel-Streiber C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA Kohl M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011301; CAC19891.1; -.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR001950; TIF_SUI1.
DR Pfam; PF01473; CW binding 1; 17.
DR PROSITE; PS01118; SUI1.1; 1.
SQ SEQUENCE 2367 AA; 269072 MW; BF7087C92E7388E7 CRC64;

Query Match 78.8%; Score 4090.5; DB 2; Length 2367;
Best Local Similarity 77.2%; Pred. No. 1.1e-152;
Matches 788; Conservative 113; Mismatches 119; Indels 1; Gaps 1;

QY 1 MNLVNAQLOKQVYVFRIOEDYVAILNALBEYHNMSSESVVEKYKLKLDINNTDNYL 60
DB 1 MSLVNRKQLEKNAVFRVQEDYVAILDALBEYHNMSNTVVEKYKLKLDINSLTDYI 60
QY 61 NTYKSGRNKALKKFKXYLTMEVLEKNSLTPVEKNLHFHWIGQINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKXYLTVEILEKNSLTPVEKNLHFHWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKPYDSNAFLINTLTKTIVESATNTNLTESFRENLDNDFDYNKFKYKMEIY 180
DB 121 VNSDYTVKPYDSNAFLINTLTKTIVESATNTNLTESFRENLDNDFDYNKFKYKMEIY 180
QY 181 DKQKHFDYKSOIENPFIIDNIITKYLSEYKDLALNKYIEESLNKLTANNNDI 240
DB 181 DKQKFINYKAAQKEENPDLDIIDDIVTYLSNEYSKOIDLNAYIEESLNKVTENSGNDV 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300
DB 241 RNFEFKTGEVFNLYQEQLVERWNLGASDILRVAILKNGVYLDVNLPGIHPDLPKD 300
QY 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVORFESALSSKSKSEI 359
DB 301 INKPDVKTAVDWEEMQLEAIMKYKEYIPEVTSKHFDTLDEEVQSSFESVLASKSKSEI 360
QY 360 FLPLDDIKVSPLEVKIAFANNSVINOALISLKDSYCSDLVINQIKNRYKILNDNLNPSIN 419
DB 361 FLPLGGIEVSPLEVKVAFAGKSIIDQALISAKDSYCSDLIKQIKNRYKILNDTLGPIIS 420
QY 420 EGTDFNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYQ 479
DB 421 QGNDFTNTNFGESLGAINEENISFIAKISYLRVGFYPEANTITLSGPTIYAGAYK 480
QY 480 DLLMFKDNSTNIHLEPELNFPPTKIISQTEQITSLWSFNQARAKSQFBEYKKGYP 539
DB 481 DLLTFKEWISIDTSLSSLRNFPFPKVNISOATEQEKNSLWQFNEERAKIQPEEYKKNYF 540
QY 540 EGALGEDNDLPQNTVLDKDYVSKLSSMKTRNKEYIHYIYVLOQDKLSYEASCNLPS 599
DB 541 EGALGEDNDLPQNTVTDKYLEKISSTKSSERGYYVHYIVQ-LQGDKISYEAAACNLFA 600
QY 600 KDPYSSILYQKNIEGSETAYYYVADAEIKEDKIRIPYQISNKRNIKLTFIGHGKSEPN 659
```

```
Db 601 KNPYDSILFQKNIEDSEVAYYNPTDSBIEQIDKYRIPDRISDRPKIKLTLIGHGKAEFN 660
Qy 660 TDTFANLVDVSLSSSIEIETILAKADISPKYIEINLLGCNMFPSYIYAEETYPGKLLKI 719
Db 661 TDIFAGLDVDSLSSSIEIETILAKADISPKSIEINLLGCNMFPSYVNVVEETYPGKLLRV 720
Qy 720 KDRVSELMPISQDSITVTSANQYEVRIINEEGKREILDHSGKWINKKEESIIDKISSKEYIS 779
Db 721 KDKVSELMPISQDSITVTSANQYEVRIINEEGKREILDHSGKWINKKEESIIDKISSKEYIS 780
Qy 780 FNPENKNIIVKSKYLHELSTLLOEIRNNANSSDIIDLEKKVMTLCEIINVASNIDQIVVEG 839
Db 781 FNPENKNIIVKSKYLPSTLLOEIRNNANSSDIIDLEKKVMTLCEIINVASNIDQIVVEE 840
Qy 840 RIEEAKNLTSDSINYIKNEFKLIESISDLYDLKHONGLDSDHSPFISFEDISKTENGFRIR 899
Db 841 RIEEAKNLTSDSINYIKNEFKLIESISDLYDLKHONGLDSDHSPFISFEDISKTDEGFSIR 900
Qy 900 FINKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEHVNTL 959
Db 901 FIDKETGESIFVETEKAFSEYANHITSEISKDTIFDTVNGKLVKKNLDAAEHVNTL 960
Qy 960 NSAPFIQSILIEVNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVVELVSTALDETI 1019
Db 961 NAAFFIQSLIGVNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDETI 1020
Qy 1020 D 1020
Db 1021 D 1021
```

RESULT 3

```
ID Q9F931 PRELIMINARY; PRT; 2367 AA.
AC Q9F931;
DT 01-WAR-2001 (TReMBLrel. 16, Created)
DT 01-WAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-WAR-2003 (TReMBLrel. 23, Last annotation update)
DE Cytotoxin B.
GN TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]_TaxID=1496;
RP SEQUENCE FROM N.A.
RC STRAIN=5340;
RX MEDLINE=20448897; PubMed=10992443;
RA Sambol S.P., Merrigan M.M., Lyster D., Gerding D.N., Johnson S.;
RT "Toxin gene analysis of a variant strain of clostridium difficile that
RT causes human clinical disease.";
RL Infect. Immun. 68:5480-5487(2000).
DR EMBL; AF217292; AAG18011.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR001950; TIE_SUI1.
DR Pfam; PF01473; CW_binding_1; 18.
DR PROSITE; PS01118; SUI1.1; 1.
SQ SEQUENCE 2367 AA; 269337 MW; D5EE715B5BD41E2F CRC64;
```

Query Match 78.6%; Score 4081.5; DB 2; Length 2367;
Best Local Similarity 76.8%; Pred. No. 2.4e-152;
Matches 784; Conservative 116; Mismatches 120; Indels 1; Gaps 1;

```
Qy 1 MNLVNAQIQKVVYKFRIQEYVAILNALBEYHNMSSESVVEKYKLKDINLNTDNYL 60
Db 1 MSLVNRKQLEKMANVRFRQEDYVAILDLEBEYHNMSSENTVEKYKLKDINSLTDTYI 60
Qy 61 NTYKSGRNKALKKPKFYLTMEVLELKNLSLTPVEKNLHFIIWIGGINDTAINYNQWKD 120
Db 61 NTYKSGRNKALKKPKFYLTMEVLELKNLSLTPVEKNLHFIIWIGGINDTAINYNQWKD 120
Qy 121 VNSDYTVKPYVDSNAFLINTLTKKTIVESATNTLTESFRENLDNPPEDYNKFKYRKMELIY 180
```

```
Db 121 VNSDYTVKPYVDSNAFLINTLTKKTIVESASNDLTLESFRENLDNPPEDYNKFKYRKMELIY 180
Qy 181 DKQKHFDYKYSOLEENPEFIIIDNIIKTYLSNEYSKOLEALNKYIEESLNTKITTANGNDI 240
Db 181 DKQKNFINNYKQAKKEENPDIIIDIVKTYLSNEYSKOIDEINAYIEESLNTKVTENSNDV 240
Qy 241 RNLEKFADEDILVRLYNQELVERWNLAAASDILRLISMLKDEGCGVYLDVILPGIQDLPKS 300
Db 241 RNPEEFKTEGVFNLYEQELVERWNLAGASDILRVAILKNIGGVYLDVDMVLCPIHPDLPKD 300
Qy 301 INKPDISI-TNTSWEMIKLEALMKYKEYIPGYTSKNFDMLEDEEVQSPSPSALSCKDKSEI 359
Db 301 INKPDVKTAVDWEEMQLEALMKYKEYIPEYTSKHFDTLDEEVQSPSPSALSCKDKSEI 360
Qy 360 FLPLDDIKVSPLEVKIAPANNSVINOALISLKDSCSLVINQIKNRYKILNDNLNPSIN 419
Db 361 FLPLGDIIEVSPLEVKIAPAKSIIINOALISAKDSCSLIIQIKNRYKILNDLTGPIS 420
Qy 420 EGTDFNTMTKIFSDKIASISNEDNMFMKITNYLKVGFAPDVRSTINLSGEGVYTGAYQ 479
Db 421 QGNDFNTMNFSGESLGAINEENISFIAKISYLRVGFYPEANTTVTLSGPTIYAGAYK 480
Qy 480 DLLMPKDNSTNHLLEPELRNPEFPKTKISQTEOEITSLWSFNQARAKSQPEEYKGYF 539
Db 481 DLLTPEKMSIDTSLSSSELNPEFPKKNISQATEQEKNSLWQFNERRAKIQPEEYKKNYF 540
Qy 540 EGALGEDNDLPAQNTVLDKDYVSKKILSSMKTRNKEYIHYIVOLQOGDKISYEASCNLPS 599
Db 541 EGALGEDNDLPSQNTVTDKEYLLEKISSSTKSSRGVYHYIVOLQOGDKISYEACNLPA 600
Qy 600 KDPYSSILYQKNIEGSETAYYVADAEIKEDKYRIPYQISNKRNIKLTFTIGHGKSEFN 659
Db 601 KNPYDSILFQKNIEDSEVAYYNPTDSBIEQIDKYRIPDRISDRPKIKLTLIGHGKAEFN 660
Qy 660 TDTFANLVDVSLSSSIEIETILAKADISPKYIEINLLGCNMFPSYIYAEETYPGKLLKI 719
Db 661 TDIFAGLDVDSLSSSIEIETILAKADISPKSIEINLLGCNMFPSYVNVVEETYPGKLLRV 720
Qy 720 KDRVSELMPISQDSITVTSANQYEVRIINEEGKREILDHSGKWINKKEESIIDKISSKEYIS 779
Db 721 KDKVSELMPISQDSITVTSANQYEVRIINEEGKREILDHSGKWINKKEESIIDKISSKEYIS 780
Qy 780 FNPENKNIIVKSKYLHELSTLLOEIRNNANSSDIIDLEKKVMTLCEIINVASNIDQIVVEG 839
Db 781 FNPENKNIIVKSKYLPSTLLOEIRNNANSSDIIDLEKKVMTLCEIINVASNIDQIVVEE 840
Qy 840 RIEEAKNLTSDSINYIKNEFKLIESISDLYDLKHONGLDSDHSPFISFEDISKTENGFRIR 899
Db 841 RIEEAKNLTSDSINYIKNEFKLIESISDLYDLKHONGLDSDHSPFISFEDISKTDEGFSIR 900
Qy 900 FINKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEHVNTL 959
Db 901 FINKETGESIFVETEKTFSEYANHITSEISKDTIFDTVNGKLVKKNLDAAEHVNTL 960
Qy 960 NSAPFIQSILIEVNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVVELVSTALDETI 1019
Db 961 NAAFFIQSLIEVNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDETI 1020
Qy 1020 D 1020
Db 1021 D 1021
```

RESULT 4

```
Q46034
ID Q46034 PRELIMINARY; PRT; 2367 AA.
AC Q46034;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Toxin B.
OS Clostridium difficile.
```



```

QY 301 INKPSI-TNTSWEMIKLBAMKYKEYIPGYTSKNFMDLDERVQSFESALSSKSKSEI 359
Db 301 INKPSVKTAVDEEEMQLEAIIMKYKEYIPEYTSKHPDTLDEEVQSFESVLASKNKSRI 360
QY 360 FLPLDDIKVSPLEVKIAPANNVINOALISLSDSYCSDLVINQIKRKYKILNDNLNPSIN 419
Db 361 FLPLGDIEVSPLEVKIAPAKGSIINOALISAKDSYCSDLIKQIQNRKYKTLNDTLGPIIS 420
QY 420 EGTDFNTWKIIPSDKLASISNEDNMFMKITYNLKVGFPADVRSITNLSGPGVYTGAQO 479
Db 421 QGNDFTNTWNGESLGAITANENTISPIAKISYLAVGYPENATITILSGTIYAGAYK 480
QY 480 DLLMFKONSTNTHLLEPELNFEPFKTKISQTEQBITSLMSFNQARAKSQPEYKKGYP 539
Db 481 DLLTPKMSIDTSILSSLELNFEPFKVNISQTEQEKSLWQFNBERAKIQPEEYKKNYP 540
QY 540 EGALGEDDNLDFQAQ 553
Db 541 EGALGEDDNLDFSQ 554

RESULT 6
O68653 PRELIMINARY; PRT; 698 AA.
AC O68653;
DT 01-AUG-1998 (TremBLrel. 07, Created)
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Truncated toxin A.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]_TaxID=1496;
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 20309;
RX MEDLINE=99304497; PubMed=10376209;
RA Song K.P., Bai X.L., Chang S.Y.;
RT "Nucleotide and peptide sequences of the open reading frame encoding a
RT truncated toxin A gene of Clostridium difficile strain CCUG 20309.";
RL DNA Seq. 10:93-96(1999).
DR EMBL; AF053400; AAC08437.1; -.
SQ SEQUENCE 698 AA; 80682 MW; 021BB268A3BDEC5E CRC64;

Query Match 36.8%; Score 1911; DB 2; Length 698;
Best Local Similarity 52.9%; Pred. No. 1.1e-67;
Matches 371; Conservative 136; Mismatches 186; Indels 8; Gaps 5;

QY 1 MNLVKAQIQKVVYKFRIOEYVAIINALBEYHNMSSESVVEKYKLKLDINNLTNDYL 60
Db 1 MSLISKEELIKLAY-SIRPRENEYKTVLTNLDYNNKLTNNNNENKYLQKLKLNESIDVFM 59
QY 61 NTYKSGRNKALKKPKFKEYLTMEVLEKNSLTPVEKQLHFIWIGQINDTAINYNQWKD 120
Db 60 NKYKSSRNRLSNLKKDILKEVILKNSNTSPVEKNLHFVWIGVSDVILEYIKQWAD 119
QY 121 VNSDYTVKPYDSNAPLINTLKTIVESATNTLTLESFRENLDNPDPEYKFKRMEIY 180
Db 120 INAEYNIKLWYDEAFNLTKAIVESSTTALQLEEEIQNPQPDNNKFKYKRMFIY 179
QY 181 DQKHFIDYKSOIEBENPFIIDNIIKTVLSNEYSKDLEALNKYIESLNKITTANNNDI 240
Db 180 DRQKRFINYKQKQKPTVLTDDIISKVLYSEYNRDETLLSEYRTNSLRKINSNGIDV 239
QY 241 RNLEKFADEDLVRLYNQELVERNLAAASDIIRISMLKEDGGVYLDVILPGIQDLPFKS 300
Db 240 RANSLFTEQELLNIYNQELNRCNLAAASDIVELLALKNFGVGYLDVMDLPGHSLDFT 299
QY 301 INKPSITNTSWEMIKLEIMKYKEYIPGYTSKNFMDLDERVQSFESALSSKSKSEIF 360
Db 300 ISRPSIGIDRWEMIKLEIMKYKEYINNTYSENFDKLPQQLKDNPKLIIESKSEKSEIF 359
QY 361 LPLLDDIKVSPLEVKIAPANNVINOALISLSDSYCSDLVINQIKRKYKILNDNLNPSINE 420

```

```

Db 360 SKLENLAVSDLEIKIAFALGVSINOALISKGSYLTNLVIEQIKRKYQFLNQHLPALIES 419
QY 421 GTDENTWTKIIPSDKLASISNEDNMFMKITYNLKVGFPADVRSITNLSGPGVYTGAQO 480
Db 420 DNNFTDTTKIIPHDSLFNSATAENSMTLKIAPIYLVQGFMEPEARSTISLSPGAYASAYD 479
QY 481 LLMPKDNSTNTHLLEPELNFEPFKTKISQTEQBITSLMSFNQARAKSQPEYKKGYP 540
Db 480 FINLQDNTIEKTLKASDLIEFKFPENNLSQLTEQEIINSLSWPDQASAKYQFEKYKVDYTG 539
QY 541 GALGEDDNLDFQAQNTVLKDY-VSKILSS--MKTRNKEYIHVIYVQLQGDKISYRASCML 597
Db 540 GSLSEDNEVDENKNTALDKYLLNKNKIPSNVVEEAGSKNYHYIIQLQGGDDISYEATCNL 599
QY 598 FSKDPYSSILVQKNTGSETAYVYVAD--ARIKIDKYRIPYQISNKNIKLTPFTGHCK 655
Db 600 FSKNPNKSIIRQNM--NESAKSYFLSDGGSILSNKYRIPERLKNKEKVKVTFIGHCK 657
QY 656 SEFNFTDTTANLDDVDSLSSEIETILNLAKADISPKYIEINLL 696
Db 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLIDISPKNVEVNULL 698

RESULT 7
Q9EXQ8 PRELIMINARY; PRT; 698 AA.
AC Q9EXQ8;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE Truncated toxin A.
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]_TaxID=1496;
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA von Eichel-Streiber C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011301; CAC19893.1; -.
SQ SEQUENCE 698 AA; 80799 MW; 66D9B21798D314BB CRC64;

Query Match 36.5%; Score 1893; DB 2; Length 698;
Best Local Similarity 52.4%; Pred. No. 5.4e-67;
Matches 367; Conservative 138; Mismatches 188; Indels 8; Gaps 5;

QY 1 MNLVKAQIQKVVYKFRIOEYVAIINALBEYHNMSSESVVEKYKLKLDINNLTNDYL 60
Db 1 MSLISKEELIKLAY-SIRPRENEYKTVLTNLDYNNKLTNNNNENKYLQKLKLNESIDVFM 59
QY 61 NTYKSGRNKALKKPKFKEYLTMEVLEKNSLTPVEKQLHFIWIGQINDTAINYNQWKD 120
Db 60 NKYKSSRNRLSNLKKDILKEVILKNSNTSPVEKNLHFVWIGVSDVILEYIKQWAD 119
QY 121 VNSDYTVKPYDSNAPLINTLKTIVESATNTLTLESFRENLDNPDPEYKFKRMEIY 180
Db 120 INAEYNIKLWYDEAFNLTKAIVESSTTALQLEEEIQNPQPDNNKFKYKRMFIY 179
QY 181 DQKHFIDYKSOIEBENPFIIDNIIKTVLSNEYSKDLEALNKYIESLNKITTANNNDI 240
Db 180 DRQKRFINYKQKQKPTVLTDDIISKVLYSEYNRDETLLSEYRTNSLRKINSNGIDV 239
QY 241 RNLEKFADEDLVRLYNQELVERNLAAASDIIRISMLKEDGGVYLDVILPGIQDLPFKS 300

```

Db 240 RANSLFTEQELLNIYNQELNLRGNLAASDIVRLLALKNFQGVYLDVDMPLGHSDFLFT 299
 QY 301 INKPSDITWTSWEMIKLEAIMKYKEYIPGVTSKNFDMLDEEVQSPESALSCKSEIF 360
 Db 300 ISRPSSIGLDRWEMIKLEAIMKYKEYINNTYSENFQDLQDLKDNFKLLIESKSEKCEIF 359
 QY 361 LPDIDIKVPLEVIAFANNVINOALISLSDSYCSDLVINQIKRYKILNDNLNPSINE 420
 Db 360 SKLENLVSDLEIKIRFALGVSINOALISKGSYLTNLVIEQIKRYQFLNQHLPATES 419
 QY 421 GTDFTNTMIFSKLASINEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYOD 480
 Db 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQVGFPEARSTISLSGPGAYASYD 479
 QY 481 LLMFKDSTNIHLLPELRFNPFPPKTKISQLTQEQEITSLWSFNOARAKSQFEYKKGYPE 540
 Db 480 FINLQENTIEKTLKASDLIEFKFPENNLSQLTQEQEINSLWSFDQASAKYQFEKYKDYTG 539
 QY 541 GALGEDNDLQFAONTVLDKDY-VSKKILSS--MKTRNKEYIHVIVOLQGDKISYEASCNL 597
 Db 540 GSLSENEVDFNKNTALDKNLLNKKIPSNVVEAGSKNYVHYIIQLQDDDISYEATCNL 599
 QY 598 FSKOPYSSILYQKNIQEGSETAYYYVAD--AEIKEIDKYRIPYQISNKNRIKLTFTGHOK 655
 Db 600 FSKNPKNSIIQRNM--NESAKSHFLSDGSEISLELNKYRIPERLKNKEKVKVTFIGHCK 657
 QY 656 SEENTDTFANLVDLSLSSEIETILNLAADISPKYIEINLL 696
 Db 658 DEFTSEPARLSVDSLSNEISSFLDTIKLDISPKNVEVNL 698

RESULT 8

O86141 ID O86141 PRELIMINARY; PRT; 697 AA.
 AC O86141;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE TcdA protein (Fragment).
 GN TCDA.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1496;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8864;
 RA Soehn F., Leukel P., Weidmann M., Eichel-Streiber C.V., Braun V.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8864;
 RA von Eichel-Streiber C.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y10689; CAA71690.1; -.
 FT NON TER 1
 SQ SEQUENCE 697 AA; 80680 MW; CD650D3372D65B73 CRC64;

Query Match 36.3%; Score 1887; DB 2; Length 697;
 Best Local Similarity 52.4%; Pred. No. 9.3e-67;
 Matches 366; Conservative 137; Mismatches 188; Indels 8; Gaps 5;

QY 3 LVNKAQLQKMYVYKFRIOEDYVAIILNALLEYHNMSESSVVEKYKLKDNLTNYLT 62
 Db 2 LISKEELIKLAY-SIRPRENEYKTLTNLDEYNKLTNNNNENKYQLKKNESIDVFMNK 60
 QY 63 YKSGRKNALKKPKFVLTMEVLEKNNSLTPVEKNLHFWIGGQINDTAINYNQWKN 122
 Db 61 YKNSRRNALNLUKKDILKEVILIKNSNTPVEKNLHFWIGGEVSDIVLEYIKQWADIN 120
 QY 123 SDYTVKFDVSDNAFLNTLTKTTIVESATNTLSPRENLDNPDYKFKRMEIYDK 182
 Db 121 AYNIRLWYDEAPLVNTLKAIVESSTTALQLLEEEIQNPQFDNMKFKRMEIYDR 180

QY 183 QKHFDYKSOIEENPEFIIIDIIKTYLSNEYSKOLEALNKYIEBSLNKITANNNGDIRN 242
 Db 181 QKRFINYKSOINKPFTVITDDIIKSYLISEYNRDETILLESYRTNSLKINSHGIDVRA 240
 QY 243 LEKFAEDLVRLYNQELVERWNLAASDILRISMLKEDGGYVLDVILPGIOPDLFKGIN 302
 Db 241 NSLFTQEELLNIYNQELNLRGNLAASDIVRLLALKNFQGVYLDVDMPLGHSDFLFTIS 300
 QY 303 KPDSITNTSWEMIKLEAIMKYKEYIPGVTSKNFDMLDEEVQSPESALSCKSEIPLP 362
 Db 301 RPSSTIGLDRWEMIKLEAIMKYKEYINNTYSENFQDLQDLKDNFKLLIESKSEKCEIFSK 360
 QY 363 LDDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINQIKRYKILNDNLNPSINEGT 422
 Db 361 LENLVSDLEIKIRFALGVSINOALISKGSYLTNLVIEQIKRYQFLNQHLPATESDN 420
 QY 423 DFTNTMIFSKLASINEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYODLL 482
 Db 421 NFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQVGFPEARSTISLSGPGAYASYDPI 480
 QY 483 MFKONSTNIHLLPELRFNPFPPKTKISQLTQEQEITSLWSFNOARAKSQFEYKKGYPEGA 542
 Db 481 NLQENTIEKTLKASDLIEFKFPENNLSQLTQEQEINSLWSFDQASAKYQFEKYKDYTGGS 540
 QY 543 LGEDDNDLQFAONTVLDKDY-VSKKILSS--MKTRNKEYIHVIVOLQGDKISYEASCNLFS 599
 Db 541 LSEDNEVDFNKNTALDKNLLNKKIPSNVVEAGSKNYVHYIIQLQDDDISYEATCNLFS 600
 QY 600 KOPYSSILYQKNIQEGSETAYYYVAD--AEIKEIDKYRIPYQISNKNRIKLTFTGHKSE 657
 Db 601 KNPKNNSIIQRNM--NESAKSHFLSDGSEISLELNKYRIPERLKNKEKVKVTFIGHGKDE 658
 QY 658 FNTDTFANLVDLSLSSEIETILNLAADISPKYIEINLL 696
 Db 659 FNTSEPARLSVDSLSNEISSFLDTIKLDISPKNVEVNL 697

RESULT 9

Q9FCX5 ID Q9FCX5 PRELIMINARY; PRT; 553 AA.
 AC Q9FCX5;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Toxin A (Fragment).
 GN TCDA.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1496;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C34;
 RX MEDLINE=20402122; PubMed=10931294;
 RA Braun V., Mehlig M., Moos M., Rupnik M., Kalt B., Mahony D.,
 RA von Eichel-Streiber C.;
 RT "A chimeric ribozyme in Clostridium difficile combines features of
 RT group I introns and insertion elements."
 RL Mol. Microbiol. 36:1447-1459(2000).
 DR EMBL; AJ131844; CAC03681.1; -.
 FT NON TER 553
 SQ SEQUENCE 553 AA; 63913 MW; DDE4551A6D8C3B25 CRC64;

Query Match 28.3%; Score 1469; DB 2; Length 553;
 Best Local Similarity 51.7%; Pred. No. 1.9e-50;
 Matches 287; Conservative 106; Mismatches 160; Indels 2; Gaps 2;

QY 1 MNLVNAQLQKMYVYKFRIOEDYVAIILNALLEYHNMSESSVVEKYKLKDNLTNYLT 60
 Db 1 MSSISKRELIKLAY-SVRPRENEYKTLTNLDEYNKLTNNNNENKYQLKKNESIDVFM 59
 QY 61 NTYKSGRKNALKKPKFVLTMEVLEKNNSLTPVEKNLHFWIGGQINDTAINYNQWKN 120


```

Db 60 NRYKSSRRNALSNLKKOITLKEVILIKNSNTSPVEKNLHFMVIGVEVSDIALEYIKQWAD 119
QY 121 VNSDYTKVPVYDSNAFLINTLTKTIVESATNTNLESFRENLANDEPDYNNKYRKRMEIY 180
Db 120 INAEYVKKLWYDSEAPVNLTKAIVESSTTALQLEEEIQNPQDM-KFYKQRMFIY 178
QY 181 DKQKHFDIYKSOIBENPEFIIDNIKIYLSNEYSKDLKALNKYIEESLNKITYANNGNDI 240
Db 179 DRQKRFINYKSOINKPTVPTIDDIKSHLVSEYNRDETLLSEYRTNSLRKINSNHGDI 238
QY 241 RNLEKPADDELVRNLQNELVERWNLAAASDIIRISMLKEDGGVYLDVILPGIQPDLFKS 300
Db 239 RANSLFTEOELLNIYSQELNLRGNLAAASDIIRLLALKNFGVYLDVILPGHSLFKT 298
QY 301 INKPDISTNTSWEMIKLEAIMKYKEYIPGYTSKNPDMLEEVQVSPESALSSKSKSEIF 360
Db 299 IPRPSIGLDRWEMIKLEAIMKYKYNNTYBENFDKLDQQKDNFKLIESKSEIF 358
QY 361 LPLDDIKVSPLEVKIAPANNVINQALISLKDSCSYDLVINQIKRYKILNDNLNPSINE 420
Db 359 SKLENLWSDLEIKIAPALGVSINQALISKQSYLTNLVIEQVKNRYQFNLQHLNPAIES 418
QY 421 GTDFNTTKIFSDKLASINEDNMFMKIITNYLVKVPAPDVRSTINLSGPGVYTGAYOD 480
Db 419 DNNFTDTTKIPHDSLFNSATAENSMLTKIAPYLVQGFMPPEARSTISLSGPGAYASAYD 478
QY 481 LLMFQDNSTNIHLLEPELRFNFPKTKISQLEQETISLWSNOARAKSQFEYKKGYPE 540
Db 479 FINLQNTIEKTLKASDLLEFPENNLSQLTEQELNSLWSFDQASAKYQIBRYVRYTG 538
QY 541 GALGEDDNLDPQNT 555
Db 539 GSPSGDNGVDNFKNT 553

```

RESULT 10

```

Q46149 ID Q46149 PRELIMINARY; PRT; 2178 AA.
AC Q46149; Q46147; Q46148;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Alpha-toxin.
OS Clostridium novyi.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxId=1542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19402;
RX MEDLINE=95342160; PubMed=7616958;
RA Hofmann F., Herrmann A., Habermann E., von Eichel-Streiber C.;
RT "Sequencing and analysis of the gene encoding the alpha-toxin of
RT Clostridium novyi proves its homology to toxins A and B of Clostridium
RT difficile.";
RL Mol. Gen. Genet. 247:670-679(1995).
RN [2]
RP SEQUENCE OF 1204-2178 FROM N.A.
RC STRAIN=ATCC19402;
RA Hofmann F., Habermann E., von Eichel-Streiber C.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z48636; CAA88565.1; -
DR EMBL; Z23280; CAA80818.1; -
DR EMBL; Z23281; CAA80819.1; -
DR InterPro; IPR002479; CW binding.
DR Pfam; PF01473; CW binding 1; 10.
SQ SEQUENCE 2178 AA; 250134 MW; 9B0ADCE031C4A75A CRC64;

```

Query Match 27.4%; Score 1424.5; DB 2; Length 2178;
 Best Local Similarity 32.7%; Pred. No. 4.8e-48;
 Matches 342; Conservative 211; Mismatches 437; Indels 55; Gaps 19;

```

QY 3 LVNKAQLQOMVYKFKRIOEDYVAILNALAEEYH-NMSESSVVEKYKLKDKINNTDNYLN 61
Db 2 LITREQLMKIASIPLKRKEPEYNILDALENFRDIEGTSVKBYISKLSKLDELVDNYQT 61
QY 62 TYKSGRNKALKKPKPEYLTMEVLEKNSLTP--VEKHLHFTWIGGOINDTAINYNQWK 119
Db 62 KYPSGGRNALCNPFDSLYSELRELTKNSRTSTIASKNSLFTWIGGPISQDSLEYNNMWK 121
QY 120 DVNSDYTKVPVYDSNAFLINTLTKTIVESATNTNLESFRENLANDEPDYNNKYRKRMEIY 179
Db 122 MENKYNIRLFDKNSLNLVNTLTKTAIQESSKVIIEQNSNILDGTGYNKPKYSRDMKLI 181
QY 180 YDKQKHFDIYKSOIEENPEFIIDNIKIYLSNEYSKDLKALNKYIEESLNKITYANNGND 239
Db 182 YRYKRELKMLYENMKQNS---VDDIIINFLNPKYDIGKLNQKNNNNKMAIGATD 238
QY 240 INLEKPADDELVRNLQNELVERWNLAAASDIIRISMLKEDGGVYLDVILPGIQPDLFK 299
Db 239 I-NTENILTNLTKSYQELIQTNLAAASDIIRAILKKGYGVYLDVILPGVNLSLFN 297
QY 300 SINKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNPDMLEEVQVSPESALSSKSKSEI 359
Db 298 DISKNGHDSNYWEAAIPEATANEKLMNNYPKYMEQVPSEIKERILSFVRN-HDINDL 356
QY 360 FLPLDDIKVSPLEV-----KIAFANNVINQALISLKDSCSYDLVINQIKRYKILNDNL 414
Db 357 ILPLGDIKISQLEILLRLKAATGKTFNAPFISNNDSLTLNLLISQLENRYEILNSII 416
QY 415 NP--SINEGTD--FNTTMMKIFSDKLASINEDNMFMKIITNYLVKVPAPDVRSTINLSG 470
Db 417 QEKFKICETYSYINSVSELVLETPKNLSMDGSGFYQOIIIGYLSGFGKPEYNSTVFFSG 476
QY 471 PGVYTGAYQDILMFKN-----STNIHLLEPELRFNFPKTKISQLEQETISLWSFPQ 524
Db 477 PNIYSATCDTYHFINKTFDMLSSQOEIFEAS--NNLYFSK-----THDEFKSWLLRS 529
QY 525 ARAKSQFEYKGYFEGALGEDDNLDPQNTVLDKDYVSKKI--LSSMKTRNKEYIHYIV 582
Db 530 NIAEKEFOKLKITYIGRTLNVYEDGLNFNKKRVTTSELLKVIIEVNSTKIYENYDLNML 589
QY 593 QLOGDKISYEACNLFSKDPYSSILYQKNIEG-SETAYYYYVADAEIKEDIKRYPQIS 641
Db 590 QIQGDDISYESAVNVFGKNPNKSI-----IQGVDDFANVYFENGIVQSDNNINLSRFN 645
QY 642 NKRNIKLTFIGHGKSEPTDTPANLVDVLSSEI-----ETILNLAKADISPKYTEINLL 696
Db 646 DIKKIKLTLIHGENVFNPKLPGKTVNDLYTNIIKPKQLHLEREGVILKNKYLKINIL 705
QY 697 GCNMFYSIYABETYPGKILLKIKDRVSELMP-SISQDSITVSANQYEVRIINEEGREIL 755
Db 706 GCMFTPKVDINSTFVGKLFNKIS---RDLQPKGFSKQLEISANKYAIRINREGREVL 762
QY 756 DHSGWKINKESIINKDISKEYISPNKENKIIIVKSKYLHELSTLLOEIRNNANSSIDL 815
Db 763 DYFGKWSNTDLIARQISNKYVYVNEVENTLSARVEQLNKVAFAKDI-----NSIIQT 817
QY 816 EKKVMTLCEINVASNIDRQIVEGRIERAKNLTSDSI---NYIKNEFKLIESISDLYD 871
Db 818 TNNQELKQSLVNTYADLTITLYSELKEDIPELDNIQIKERIIILNEISRLHDFNSIID 877
QY 872 LKHQGLDSDHSFISFEDISKTENGFRIRPINKETGNSPIETEKEIFSFYATHISKEISN 931
Db 878 FYQKNISNNMIILFDSIIKEKDYNNVKLANKITGETSVIKTYSDSLWNFTNKYKIVDD 937
QY 932 IKDITFDNVNGLVKKNLNDAAHEVNTLNSAPFIOSLIEYNTTKESLSLSLVAMKVQVTA 991
Db 938 IKGIIVKQINGEPIKADFEBQNPSSLNSAMLQMLLDIDYKPYTEILTNTMTSLKVQAYA 997
QY 992 QLFSTGLNTITDASKVVELVSTALD 1016
Db 998 QIFQISGAIQAEITVITIISDALN 1022

```

```

RESULT 11
Q9ZGR4 PRELIMINARY; PRT; 3169 AA.
AC Q9ZGR4;
DT 01-WAY-1999 (TrEMBLrel. 10, Created)
DT 01-WAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative cytotoxin.
GN L7095
OS Escherichia coli O157:H7.
OG Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
[1] SEQUENCE FROM N.A.
RP STRAIN=EDL933;
RC MEDLINE=98391744; PubMed=9722640;
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA Blattner F.R.;
RT "The complete DNA sequence and analysis of the large virulence plasmid
of Escherichia coli O157:H7."
RL Nucleic Acids Res. 26:4196-4204 (1998).
DR EMBL; AF074613; AAC70163.1; -
DR InterPro; IPR006473; Yop.T.
DR TIGRfams; TIGR01586; yopT_cys_prot; 1.
KW Plasmid.
SQ SEQUENCE 3169 AA; 361971 MW; C2CDB48A72BD5FD1 CRC64;

Query Match 7.9%; Score 411.5; DB 2; Length 3169;
Best Local Similarity 20.9%; Pred. No. 4.3e-08;
Matches 257; Conservative 195; Mismatches 453; Indels 323; Gaps 54;

QY 14 YKFR1---QDEYVAILNAL--EYHNMSSESVVEKYKLKDINNLTNYLNTYKSKR 68
DB 160 YIKIRTRGAEDQTTITQSLIINELLNGVDRNTI--PFQKISELNDIIHSYENMQIKSR 218
QY 69 N--KALKKPKVYLTMEVLEKONS-----LTPVEKNLHFTWIGGO 106
DB 219 KGIEILVKGQELSSLINONKGNKQKSDNASKIINLLGIEYQSHKVDFPFHAWVAGA 278
QY 107 INDTAINYINQWKNVSDYTVKPYDSNAP-----LINTLAKT----- 144
DB 279 PPDNTSYITAFNTKYDYTLWIDPNAGAAKPSGILKNIAMVYAIMLRRTNPHLAE 338
QY 145 -----IVESATNTLE--SPRENLDPEFY-----NKFYKRMIEIYDQKHFI 187
DB 339 EMNEVILKIQNETIEFKETRERLKELENRYKSLTSETKGFNVFFLESIMGMQDNFY 398
QY 188 DYYKQIENPEFI-----IDNLIK--TVLSNEYSKDLEALNKYIEESLNKITTANNGN-- 238
DB 399 TYCISNGISNTDDISLDFLTNVKLKSPVQNDKSTVEKNKRDIDLLKNTISQKFGDRF 458
QY 239 ---DIRNLEKFADEDLVRLYNQBLVERWNLAAASDILRISMLKEDGGVYLDVILPGIQP 295
DB 459 QLRDINTLESFKKPQDYFFQEQEMLLRWVAAASQVRINILKEYGIYTDITLPAYS 518
QY 296 DLPKSNKPDSTNTGWEMIKLEAIMKYK--EVIPO--YTSKNFMDLDEEVORSFSAIS 351
DB 519 KVSQIINE-KSDDKRFEDLKLRIISSESILSLIKGEYSIKH--DGLDETTLNQLNML- 575
QY 352 SKSDKSEIFLPLDDIKVSPLEVKIAPANNSVI-----NOALISLKDYSYC 395
DB 576 SEIEK-----LTIDY--FKPVETKVRDTPFKIFRYQKWTENTWIRGNMFMTHKGSKC 630
QY 396 SDLVINQIKNRYKILNDLNPNISNEGDFNTTWKIFSDKLASTISN-----DNM 444
DB 631 IDFILSQKQKYLEL-QRIRDNISYNNLFVTT-----EDLKSINNVAGIGIPAKKYLEHG 684
QY 445 MFMIKITYLKVGAFDPVRSNTINLSP----- 471
DB 685 LF-----SEYRQDGTPIYVSTLNSISGPDIMTQMVKYKSLGRIGEVHIKDNKLSVDNFL 740

```

DR TIGRFAMS; TIGR01586; yopT cys prot; 1.

KW	Plasmid.
SQ	SEQUENCE 3169 AA; 361943 MW; D9BED726A62DDBFF CRC64;

Query Match 7.9%; Score 408.5; DB 2; Length 3169;

Best Local Similarity	20.7%; Pred. No. 5.7e-08;	
Matches 255; Conservative	200; Mismatches 449;	Indels 325; Gaps 55;

```

Qy 14 YVKFRI---QEDEYVAILNAL--BEYHNMSSESVVEKYKLKDINLNTDNYLNTYKSGR 68
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 160 VIKIRKTRGAEDQPTTTTQSLINELNGLVDNRNTI-PPQKISLENDIHSYENMQIKNSR 218

```

Qy	69	---	NKALKKFK	----	YLTMEVLEKNSLSLTPVEKNLHPITGG	105
Db	219	KGIEILVKQ	GELLSSLLN	VGNKGNKQLSD	NASKIINLIGIEYQSHKV-DIEPFTHAVVWAG	277

Qy	106 QINDTAINYNOWKDVSNDYTKVFVDSNAF-----LINTLKKT-----144
Db	278 APPDNTSEYITAFNLNTKDYTYLLWIDFNAGFAAKFGSGLKNIAMNVAIMRLRNTPHLA 337

Qy	145	-----	IVESATNTLE--	SFRENLDPEFY	-----	NKPYRKMEIIVDKQHF	186
D6	338	---	EEEMVILKIQNE	IEFKETRRLKE	LENRYKLS	TSKTKFNFFLES	397

Qy	187	IDYKSOIEENPEFI	----	IDNIK--	TYLSNEYSKDLEALNKYIEESLNKIKITANNGN-	238
		:	:	:	:	
		:	:	:	:	
Db	398	FTYCISNGISNTDDISRLDFTLVNLKLSPEVQNDFKSTVEKNKRDIIDLKNTISQKFGDR	457			

[illegible]

QY	295	PDLFKSINKPDSITNTSWEMIKLEAIMKYK--EYIPG--YTSKNPMDLDEEVORSFESAL	350
	:	:	:
Db	518	DKVQSIINE-KSDQKRFEDLKLRIISLSESILSLIKGKYSIKH-DGDETTTLQLNNIL	575
	:	:	:

[illegible]

Qy	395	CSDLVINQIKRYKILNDNLNPSINEGDFNTMTKFSDKLASISNE-----DN	443
D6	630	CIDFILSGQQGYLEL-QRIRDNISYNNLFYTT-----EDLSLNNAVGIGIPAKKYLEH	683

Qy	444	MMFMKITYLKVGFADVRSTINLSCP	-----	471
		: : : : :		
		: : : : :		
Db	684	GLF-----SEYRGDTIPYVSTINISGPD	MIIMRQMKKYKSLGRIGEVHIKDKLS	YVNF 739

Qy	472	GVYTGAYODLLMF---	KDNSTNIHLLPE	-----	LRNFEPKTKISOLT	-----	EQBIT	517
			:	:	:	:	:	
		:	:	:	:	:	:	
		:	:	:	:	:	:	
		:	:	:	:	:	:	
Db	740	LGVIASNNKDKNSFNWLNPV	SGVINDITPDDESSWAVRN	DINKILPFKINCHVPEKLP	799			

QY	518	SLWSFNOARAKSQFPEYKGYFEGALGEDDNLDPQNTVLDKDYVSKILSSMKTNRKEY	577
	: :	: :	
Db	800	SLY-----YEIDSRSPFGWG-----DNKSIKMHVTEINKOLI--KDINLLTSSNID	843
	: :	: :	

```

Qy      578  IHYIVQLGDKTISYEASCNLFKDPYSILYQKNIEGSETAYYYYVADAEIKDKYTRIP 637
      844  VKLLIKL--DRELYAISKI-----DNPLAISRITLQQLAVTSTFENTINFI 895
Db

```

Qy	638	YQISNKRN-----	IKLTFIGHCKSEFNTDTFANLVDVS-----	LSSIE 676
Db	896	YDFYRKQDLLLLSAIK-----	FSRNDATKIIWYNSYMEKNVFLREVIS 941	

Qy	677	TIINLAKADISPKYIEINLLGCNMFYSYIYAETYPGKLLLLKTKDRVSELMPSISQDSI-	735
		:: :: :: :: :: :: ::	
Db	942	CVLSRRKKVD---SYINEN-----KKNLSKEDA--GALRDYAKLKKMELSNLDDDCGYK	989

Qy	736	-TVSANCYEVNRNEEGKREILDHSGKWINKESIIKDSSKEVISP-	782
	:	:	:
Dd	990	KIITTNAY--IKERDKL-----SGIIYNIEHSIIIGSHSFDIRSNQHEWGLSTVEQF	1041

QY 783 KENKIIVKSXYLHELSTLLQEIENNANSSDIDLEKKVMLTECEINVAS-----NIDROIV 837

1042 KEEFYKSE-LSSAKSIFDIK-K-YITDETGRNLYHOJSDPKERIAFDISHYAY 1099

QY 838 EGRIEAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHQNG 877

[illegible]

QY 935 TIFDNVNGKLKKVKNLDAAEHVNTLSAFFIQSILIEYNTTK-----ESLSNLSV---- 983

DB 1207 QVEMKTEQSSVIGIKVSVNEDPVNNTFLSIPINONTIDEMVSVEMIVNHTCTE 1269

QY	984	--AMKVOVYAQLFSTGLNTITDASKVVVL	1010
		: : : : : : : : : : : :	
DB	1267	NVNPSTVYSSSTNIDETINDPQVINK	1295

RESULT 13
C98119

ID	Q8L1L9	PRELIMINARY;	PRT; 3223 AA.
AC	Q8L1L9;		
DT	01-OCT-2002	(TTEMBLrel. 22, Created)	
DT	01-OCT-2002	(TTEMBLrel. 22, Last sequence update)	

01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DT
DE
DE
EFal-Lymphostatin-like protein.
OS
OC
CC

OC Enterobacteriaceae; Escherichia.
OX
RN [1]
PP NCBI_TaxID=562;
SEQUENCE FROM N A

RC	STRAIN=493/89;
RA	Janka A.;
RT	"Identification and Distribution of the Enterohemorrhagic Escherichia
PT	coli Factor for Adherence (efal) Gene in Sorbitol Fermenting

RT Escherichia coli O157:H-";
RL Thesis (2002), Department of Institut fuer Hygiene,
RL University of Muenster, Muenster, Germany.
DE EMBL: AJ459584. CAD10848.1. --

```

DR DR InterPro: IPR001917; NHTransf_2.
DR DR InterPro: IPR001169; SHProt_acsite.
DR DR InterPro: IPR006473; Yop_T.
DR DR TIGRfams: TIGR01586; yopT_cvs_prot. 1
DR DR TIGRfams: TIGR01586; yopT_cvs_prot. 1

```

DR	PROSITE; PS00599; AA TRANSFER CLASS 2; 1.
DR	PROSITE; PS00639; THIOLE PROTEASE HIS; 1.
SQ	SEQUENCE 3223 AA; 365885 MW; _LA65A3B0A7C6E6ED CRC64;

Query Match	7.4%	Score 386;	DB 2;	Length 3223;
Best Local Similarity	19.9%	Pred. No. 4.4e-07;		
Matches 255;	Conservative 203;	Mismatches 432;	Indels 392;	Gaps 55;

Qy 2 NLVNKAOLQKMYVVFRIQEDBYVAILNALBSEYHNHSS-----VVEKYLK-LKIDNN 54
Db 226 NFKVQISLSDELTKKY--ANEILEIKRIMGEYNLLPDKNSRNGLKLLQKQADLLKIME 282

Qy	55	LTDNYLTYKSGRNLKPKEPYLEVLELN---	NSLTPEVKNLHFITWGGQNDA	111
				:
Dd	283	DTSVTENTFKN-----	IEMAIADIKREYYSHTVTVDIEKNTHAIWAGSPPSI	329

Qy	112	111	110	109	108	107	106	105	104	103	102	101	100	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79	78	77	76	75	74	73	72	71	70	69	68	67	66	65	64	63	62	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0																																																																																																																																																																																																																										
Db	330	329	328	327	326	325	324	323	322	321	320	319	318	317	316	315	314	313	312	311	310	309	308	307	306	305	304	303	302	301	300	299	298	297	296	295	294	293	292	291	290	289	288	287	286	285	284	283	282	281	280	279	278	277	276	275	274	273	272	271	270	269	268	267	266	265	264	263	262	261	260	259	258	257	256	255	254	253	252	251	250	249	248	247	246	245	244	243	242	241	240	239	238	237	236	235	234	233	232	231	230	229	228	227	226	225	224	223	222	221	220	219	218	217	216	215	214	213	212	211	210	209	208	207	206	205	204	203	202	201	200	199	198	197	196	195	194	193	192	191	190	189	188	187	186	185	184	183	182	181	180	179	178	177	176	175	174	173	172	171	170	169	168	167	166	165	164	163	162	161	160	159	158	157	156	155	154	153	152	151	150	149	148	147	146	145	144	143	142	141	140	139	138	137	136	135	134	133	132	131	130	129	128	127	126	125	124	123	122	121	120	119	118	117	116	115	114	113	112	111	110	109	108	107	106	105	104	103	102	101	100	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79	78	77	76	75	74	73	72	71	70	69	68	67	66	65	64	63	62	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0

Qy 172 Y---RKR-----METYDQKHFIDY-- 189
 ||: | | | |
Db 390 YNEIRKKNYNPSGQOEYLNKLRELATYQKISTPLKHMENSFLENNIKLQDNFFNYCI 449

```

Qy 190 YKSQIENPEFID---NIITYLSN---YSKJLEALNKYIEESL-----NKIT 233
      |||      |||      |||      |||      |||      |||      |||
Db 450 VKGVTEINDELRLNYLNQVVK--LSDDDIGNYQKTINDNKDRVKKLILDLQKQFENRIS 507

```

QY 234 ANNGDIRLEKFADEDLVRLYNQELVERWNLAAADILRISMLKEDGGVYLDVILPGI 293
DB 508 IKDVNSLTSLSKSENN---HNYQTEMLLRWNPAAASDLRLMYLKEHGGIYTDMPAY 564
QY 294 QPDLFINSKPSDITNTSWEMIKL-----EAMKYKEYIPGYSKNFMDLD-EVQORSPE 347
DB 565 SKQVIFKIMQMGNDNRFLDLKLRRAISDGLVRY-----VNNQNIDEVNYNEI----- 613
QY 348 SALSSKSKSEIFPLDDIKVSPLEVKIAPANNSV-----INQ 385
DB 614 ----SDADKNIKKILTEISKMPESIFTKINTRIPROTMTPLRLRYHLWPGWNRGLNG 669
QY 386 ALISLKSDYCSDLVI---NQIKNRYKILNDNLNPSINEGTDFNTMTKIFSDKLASINED 442
DB 670 FMLSHKSGEVIDAVIAGQNAVRELRIRDNHSEIY---FKQT-----DELSLDPD 720
QY 443 NMFMIKITNYLKVQ-----FAPDVRSTINLSGP----- 471
DB 721 KIGGIL-VKKYLSGSLFSKFRQDTIPEALSTLQISGPDLIQKMLQFRSRGVLGEBFI 779
QY 472 ----GVY-----TGAVQ-----DILMFKN 488
DB 780 NERKLSKAYIGYKTTGTCKYDMLTPESIGVNDVTPADESTWCIGKRCVDDFLFKOVS 839
QY 489 TNHILLEPELRNPEFPKTKISQLTEQIEITSLWSFNQAR-AKQFPEYKKGVEGALGEDD 547
DB 840 T-----LKTENLEPFLTKIDTTFPSQNSSTKQKOLQKIDQTLVRYNE--LIDSS 889
QY 548 NLDFQONTVLDK-----DYVSKTILSSMKTRNKEYIHVYQLOQDKISYASCNL 597
DB 890 TIDFKNLYEIDQMLHIMLEMNDIAKRSLSFLOVQIAEKIRMT-----IPVDNINI 943
QY 598 FSKOPYSSILYQKNIBGSETAYYYYVADAEIKEDKIRIPYQISNKNRNIKLTFIGHKSE 657
DB 944 YPD-----LHKQNDLWSIKGFLAS--NPHTKINILYSNKTEHNI----- 983
QY 658 FNTDTPANLDVLSSEIETILNLAKADISPKYIEINILGNCNMFSSIVAEETYPCKLL- 716
DB 984 FIKOLFS---FAMVENLEDRDINNMSKDKTP-----ENWEGRVL 1020
QY 717 ----LKIKDRVSELMPSISODSITVSANQYE---VRINEEG-KREILDHS--GKWINK 765
DB 1021 QRYLELKMKHLSQSSQEAPELEISTFIYENDFLREKIEAVKKNKNSHLYFEKIKKE 1080
QY 766 ESIIDKISKEY-----ISFNPKNKIIVSKYLHELSTLLQEIIRNNANSSDIDL 815
DB 1081 QNTQDLSTKEQKLQIKALKKEISGNTKED---SHYDRLLDAFLK--KHEN-----I 1128
QY 816 EKKVMLTECEINVAS-----NIDROIVEG---RIEEAKNLTSO-----SINY 854
DB 1129 HNKIQIKDFEKEYSVAIHNIDKVIKGTDLRLYHEGVFSDINWLSRYTLHGLGITG 1188
QY 855 IKNEFKLIESISDLYDLKHQGLDSDHSFISFEDISKTENGFRIRFINKETGNSPIETE 914
DB 1189 VHTENLLPAPSSLLINILKE-----HYNEDEISAKPLAYDY-ILNKESSSIPVEIL 1241
QY 915 KEIFS-----EYATHISKETISNKTIDFNVNGKLKVKV----- 948
DB 1242 NKLSELPPHLLTPVLGQSVPLGMGYSNDGKITQEVIVSGADGPDNPSIGLIVTYLED 1301
QY 949 --NLDAAEHVNTLSAFTQSLIEYNT-----TKESLSNLSVAMKVQVYLAFLS-----T 996
DB 1302 LYNHVRMEGTLNSQ-NLRQLLENSVSSCFLTEQISKNLLSEAEKRPYQSITIEHQHULT 1360
QY 997 GLNTITDASKVVELVSTALDET 1018
DB 1361 GLPTIADAT--LSLLSVGLPGT 1380

RESULT 14

Q9RPH1

ID Q9RPH1

PRELIMINARY; PRT; 3223 AA.

Q9RPH1;
01-MAY-2000 (TREMREL. 13, Created)
01-MAY-2000 (TREMREL. 13, Last sequence update)
01-MAY-2000 (TREMREL. 13, Last annotation update)
EHEC factor for adherence.
Escherichia coli.
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
NCBI_TaxID=562;
[1]
SEQUENCE FROM N.A.
STRAIN=E45035;
MEDLINE=20117988; PubMed=10652089;
Nicholls L., Grant T.H., Robins-Browne R.M.;
"Identification of a novel genetic locus that is required for in vitro
adhesion of a clinical isolate of enterohaemorrhagic Escherichia coli
to epithelial cells.";
Mol. Microbiol. 35:275-288(2000).
EMBL; AF159462; AAD49229.2; .
InterPro; IPR001917; NHRansf_2.
DR InterPro; IPR000169; SHprot_acsite.
DR InterPro; IPR0006473; Yop_T.
DR TIGRFAMs; TIGR01586; yopT_cys_prot; 1.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS_1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS_1.
SEQUENCE 3223 AA; 365949 MW; CAP59F69242D07A CRC64;
Query Match 7.4%; Score 384; DB 2; Length 3223;
Best Local Similarity 19.8%; Pred. No. 5.3e-07;
Matches 254; Conservative 203; Mismatches 433; Indels 392; Gaps 55;
QY 2 NLVNAQLOKQVYKFRIOEDYVAILNALSEYHNMSSESS-----VVEKYLK-LKDINN 54
DB 236 NFVQKISLSDELTKY---ANEIIBKIRMGVEYNLLPDKNSRNGKLKQKQADLLKIIME 282
QY 55 LTDVNLNTYKSGRNKALKKFKELYTMEVLEKN---NSLTPVEKNLHFIWIGQINDTA 111
DB 283 DTSVTENTFKN-----IEMAITDKREYYSHTVDIEKNIHAIWAGSPESI 329
QY 112 INYINQWQDVNSDYTVKFEVDSNAFLINTLTKKTIVESATNTNTLESFRENLPNDPEFDYK 171
DB 330 SDYIKTFLTKYKFTYLLWVDEKAFGAAKFTSLVKQIAFDLACRTIQNTPOKNDIFNL 389
QY 172 Y---RKR-----MEIYDQKHIDY-- 189
DB 390 YNEIRKKYNNPSSGQOQYLNKRLRELYATYQIKISTPLKHMFSFFLENMIKLDQNFNYCI 449
QY 190 YKSQIEENPEFIID---NIKITLSNE---YSKDEALNKYIBESL-----NKIT 233
DB 450 VKGVTIENDELRLNLYKKNVVK---LSDDDIGYQKTIINDKRVKLLIIDLQKQFGENRIS 507
QY 234 ANNGDIRLEKFADEDLVRLYNQELVERWNLAAADILRISMLKEDGGVYLDVILPGI 293
DB 508 IKDVNSLTSLSKSENN---HNYQTEMLLRWNPAAASDLRLMYLKEHGGIYTDMPAY 564
QY 294 QPDLFINSKPSDITNTSWEMIKL-----EAMKYKEYIPGYSKNFMDLD-EVQORSPE 347
DB 565 SKQVIFKIMQMGNDNRFLDLKLRRAISDGLVRY-----VNNQNIDEVNYNEI----- 613
QY 348 SALSSKSKSEIFPLDDIKVSPLEVKIAPANNSV-----INQ 385
DB 614 ----SDADKNIKKILTEISKMPESIFTKINTRIPROTMTPLRLRYHLWPGWNRGLNG 669
QY 386 ALISLKSDYCSDLVI---NQIKNRYKILNDNLNPSINEGTDFNTMTKIFSDKLASINED 442
DB 670 FMLSHKSGEVIDAVIAGQNAVRELRIRDNHSEIY---FKQT-----DELSLDPD 720
QY 443 NMFMIKITNYLKVQ-----FAPDVRSTINLSGP----- 471
DB 721 KIGGIL-VKKYLSGSLFSKFRQDTIPEALSTLQISGPDLIQKMLQFRSRGVLGEBFI 779
QY 472 ----GVY-----TGAVQ-----DILMFKN 488

```

Db 780 NERKLSKAYIGVYKTTGKYDWTLPESIGVNDVTPADESTWCIGKGRCDVDFLFKQVS 839
QY 489 TNIHLLPELRNFEPKTKISQITSEITSLWSFNQAR-AKQSFEEYKKGYPEGALGEDD 547
Db 840 T-----LKTENLPFLTKIDTDFPSQSWSTTKDQKKIQDQTLVRYNE--LIDSS 889
QY 548 NLDFQANTVLDK-----DYVSKKILSMKTRNKEYIHYIYVLOQDKISYBASCNL 597
Db 890 TIDFKNLYEIDQMLHMLMNMDDIAKRSLSQVQIAEKIRMT-----IPVDNIINI 943
QY 598 FSKDPYSSILYQKNIEGSTAYYYVADAEIKEDYRIPYQISNKRNIKLTFIGHGKSE 657
Db 944 YPD-----LHKNDNDLSMSIKGFLAS--NPHTKINILYSNKTENHNI----- 983
QY 658 FNTDFTANLVDLSLSEIETILNLAADISPKYIEINLLGCMNFSYIYAEETYPGKLL- 716
Db 984 FIKDLFS---FAVMENELRDIINNMSKDKTP-----ENWEGRVML 1020
QY 717 -----LKIKDRVSELMPSISODSITVSANOYE---VRINEEG-KREILDHS--GKWINKE 765
Db 1021 QRYLELKMKDHLSQLSQSQEANEFLSITFIYENDFLREKIEAVKKNMNSHLYFEKIKKE 1080
QY 766 ESIIDKISKEY-----ISFNPENKNIIVSKYKYLHELSTLLQEIIRNNANSSDIDL 815
Db 1081 QNTWODLSLTKQKQLIKALKESIGNTEKD---SHYDRLLDAPFK--KHNE-----I 1128
QY 816 EKKWMLTECEINVAS-----NIDROIVEG---RTEAKNLTSD-----SINY 854
Db 1129 HNKIQRIKDFEKEYSVAIHNDKVIKFGQTLDRLYHEGYVFSINTLSRYTLHLGLGTG 1188
QY 855 IKNEFKLIBSIDSYLKHONGLDSDHFSIPSEDISKTENGFRIRPINKETGNSIPETE 914
Db 1189 VTEENLLPAPASSLINILKE-----HYNEDEISAKLPLADY-ILNKKSSSIPVEIL 1241
QY 915 KEIFS-----EYATHISKEISNIKOTIFDNNVNGKLKVKV----- 948
Db 1242 NKLSELPPHELLTPVLQSVNPLGMSYSDNGKITEQVIVSGADGPDNPSIGLIVTYLED 1301
QY 949 --NLDAHEVNTLSNAFFQSLIEYNT-----TKESLNSLVAMKVQVYQAQIFS-----T 996
Db 1302 LYNIHVRMEGTLNSQ-NLRQLLENSVSSCFLTEQSINKLSEAEKRPYQSUTEIHOHLT 1360
QY 997 GLNTITDASKVVELSTALDET 1018
Db 1361 GLPTADAT--LSLLSVGLPGT 1380

RESULT 15
Q9RM48
ID Q9RM48 PRELIMINARY; PRT; 3223 AA.
AC Q9RM48;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Lymphostatin.
GN LiFA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=562;
RN [1]_TaxId=562;
RP SEQUENCE FROM N.A.
RC STRAIN=E2348/69;
RX MEDLINE=20187489; PubMed=10722613;
RA Klapproth J.M.A., Scaletsky I.C.A., McNamara B.P., Lai L.C.,
RA Maltrom C., James S.P., Sonnenberg M.S.;
RT "A large toxin from pathogenic Escherichia coli strains that inhibits
RT lymphocyte activation."
RL Infect. Immun. 68:2148-2155(2000).
DR EMBL; AJ133705; CAB55629.1; -.
DR InterPro; IPR001917; NHRtransf.2.
DR InterPro; IPR000169; SHprot_acsite.

```

```

DR InterPro; IPR006473; Yop T.
DR TIGRFAMs; TIGR01586; YOP1_Cvs_prot; 1.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
DR PROSITE; PS00639; THIOLEPROTEASE_HIS; 1.
SQ SEQUENCE 3223 AA; 365963 MW; 03E45CEDDE7938C1 CRC64;

Query Match 7.3%; Score 381; DB 2; Length 3223;
Best Local Similarity 19.7%; Pred. No. 7e-07;
Matches 253; Conservative 204; Mismatches 433; Indels 392; Gaps 55;

QY 2 NLVNAQLOKQVYKFRIOEDYVAILNALSEYHNMSESS-----VVEKYLK-LKDINN 54
Db 226 NFVQKISLSDLKTKY---ANEIIEIKRIMGEYNLLPDKNSRNGKLQKQADLLKIME 282
QY 55 LTDNYLNTYKSGRNKALKKPKYELTMEVLEKN---NSLTPVEKNLHPIWIGQINDTA 111
Db 283 DTSVTENTPKN-----IEIAITDIKREYYSHTVDIEKNIHAIWAGPPESI 329
QY 112 INYINQWKNVSDYTVKPYDSNAFLINTLTKTIIVESATNNTLESFRENLDNPEFYKFK 171
Db 330 SDYIKTFLTKYKFTYYLWVDEKAPGAAKFTSVLQIAFDLACRTIQQNTPKQKIDFINL 389
QY 172 Y---RKR-----NEIYDKQKHIDY-- 189
Db 390 YNEIRKCKYNNPSSGOQYLNKLRELYATYQISTELKMFNSPFLENMIKLQDNPFNYCI 449
QY 190 YKSOIEENPEFTID--NIITKYLSE---YSKDEALNKYIBESL-----NKIT 233
Db 450 VKGVTIEINDELRIYLNKVKIK--LSDDDIGNYQKTIINDKRVKGLIIDLQKQFENRIS 507
QY 234 ANNGNDIRNLEKFADELRLVYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPQI 293
Db 508 IKDYNLSLTSKSENN---HNYQTEMLRWNYPAASDLLRWYLKHEGGIYTTDMPAY 564
QY 294 QPDLFKSINKPDSITNTSWEMIKL-----EAIMKTYEIPGYTSQNFMDL-EVQORSFE 347
Db 565 SKQVIFKIMMQTNGNRFLEDLKLRAISDGLRY-----VNNQNIDEVNRYNEI-- 613
QY 348 SALSSEKSKSEIFLPLDDIKVSPLEVKIAPANNSV-----INQ 385
Db 614 ---SDAKNIKKITLSEIKMPEDSIPTKINTRIPROTMTPLRRYHLWPDGWNIRGLNG 669
QY 386 ALISLSDSYCSDLVT---NQIKNRYKILNDNLNPSINEGTDFNTMTKFSKDLASISND 442
Db 670 FMLS HKGSEVIDAVIAGQNAVRELRIRDNIHSEIY---PKQT-----DELSLSPD 720
QY 443 NMFMKIKTNLYKVG-----PADVRSTINLSGP----- 471
Db 721 KIGGIL-VKKYLSGLSFLSKFRQDTIIEALSTLQISGPDLIQKQMLQFFRSRGVLGEFI 779
QY 472 -----GVY---TGAYQ-----DLLMFKONS 488
Db 780 NERKLSKAYIGVYKTTGKYDWTLPESIGVNDVTPADESTWCIGKGRCDVDFLFKQVS 839
QY 489 TNIHLLPELRNFEPKTKISQITSEITSLWSFNQAR-AKQSFEEYKKGYPEGALGEDD 547
Db 840 T-----LKTENLPFLTKIDTDFPSQSWSTTKDQKKIQDQTLVRYNE--LIDSS 889
QY 548 NLDFQANTVLDK-----DYVSKKILSMKTRNKEYIHYIYVLOQDKISYBASCNL 597
Db 890 TIDFKNLYEIDQMLHMLMNMDDIAKRSLSQVQIAEKIRMT-----IPVDNIINI 943
QY 598 FSKDPYSSILYQKNIEGSTAYYYVADAEIKEDYRIPYQISNKRNIKLTFIGHGKSE 657
Db 944 YPD-----LHKNDNDLSMSIKGFLAS--NPHTKINILYSNKTENHNI----- 983
QY 658 FNTDFTANLVDLSLSEIETILNLAADISPKYIEINLLGCMNFSYIYAEETYPGKLL- 716
Db 984 FIKDLFS---FAVMENELRDIINNMSKDKTP-----ENWEGRVML 1020
QY 717 -----LKIKDRVSELMPSISODSITVSANOYE---VRINEEG-KREILDHS--GKWINKE 765
Db 1021 QRYLELKMKDHLSQLSQSQEANEFLSITFIYENDFLREKIEAVKKNMNSHLYFEKIKKE 1080

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:25:02 ; Search time 51.25 Seconds
(without alignments)
5265.077 Million cell updates/sec

Title: US-09-126-816B-6_COPY_1_1700
Perfect score: 8677
Sequence: 1 MNLVNAQLQKQVMVYKFRQ.....YLYGIDRYNVKVIAPNLYT 1700

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6730.5	77.6	2366	17 AAR95011	C. difficile toxin
2	6730.5	77.6	2366	19 AAW68388	Clostridium diffi
3	4458	51.4	2710	17 AAR95016	C. difficile toxin
4	4458	51.4	2710	19 AAW68387	Clostridium diffi
5	2780	32.0	546	20 AAY33700	C. sordellii letha
6	398	4.6	1979	21 AAB18171	Plasmodium falcipa
7	357.5	4.1	5024	22 AAG82935	S. epidermidis ope
8	348	4.0	10182	23 ABP38314	Staphylococcus epi
9	337	3.9	2485	21 AAB18172	Plasmodium falcipa

10	314.5	3.6	2496	21 AAB18222	Plasmodium falcipa
11	313	3.6	3973	21 AAB18253	Plasmodium falcipa
12	310	3.6	2184	22 AAE00425	P. falciparum telo
13	309	3.6	1639	19 AAW54145	P. falciparum synt
14	309	3.6	1639	23 AAE29345	Plasmodium falcipa
15	306.5	3.5	1516	21 AAB18195	Plasmodium falcipa
16	303.5	3.5	1802	21 AAB18217	Plasmodium falcipa
17	302	3.5	1254	11 AAR07503	Merzoite apical-en
18	302	3.5	1254	18 AAW24575	Merzoite apical-e
19	297.5	3.4	1308	21 AAB18167	Plasmodium falcipa
20	297	3.4	1654	6 AAP50777	Sequence of the Pi
21	296	3.4	2010	21 AAB18218	Plasmodium falcipa
22	293	3.4	2539	21 AAB18198	Plasmodium falcipa
23	288.5	3.3	1948	23 ABP73774	Candida albicans e
24	287.5	3.3	1844	21 AAB18250	Plasmodium falcipa
25	287	3.3	2380	21 AAB18315	Plasmodium falcipa
26	287	3.3	3433	18 AAW22017	Utrrophin. Homo sa
27	284	3.3	1881	23 ABP73809	Candida albicans e
28	282	3.3	2295	21 AAB18180	Plasmodium falcipa
29	282	3.2	2681	24 ABJ19025	Pathogen specific
30	281	3.2	1436	21 AAB18199	Plasmodium falcipa
31	281	3.2	3696	23 ABP40235	Staphylococcus epi
32	280	3.2	1817	21 AAB18301	Plasmodium falcipa
33	275	3.2	1712	21 AAB18205	Plasmodium falcipa
34	274	3.2	4134	20 AAY31946	Plasmodium falcipa
35	272	3.1	980	21 AAB18294	Plasmodium falcipa
36	271	3.1	2633	22 ABG06505	Novel human diagno
37	271	3.1	7201	22 ABB71136	Drosophila melanog
38	268.5	3.1	1847	21 AAY52002	M. jannaschii MJ14
39	268.5	3.1	1847	21 AAY51631	M. jannaschii MJ14
40	268	3.1	1477	20 AAY19981	B. burgdorferi ant
41	268	3.1	1494	20 AAY19980	B. burgdorferi ant
42	268	3.1	2663	22 AAM39097	Human polypeptide
43	267.5	3.1	2573	21 AAB18234	Plasmodium falcipa
44	267	3.1	1817	21 AAB18255	Plasmodium falcipa
45	266.5	3.1	1121	21 AAB18241	Plasmodium falcipa

ALIGNMENTS

RESULT 1
AAR95011
ID AAR95011 standard; Protein; 2366 AA.
XX
AC AAR95011;
XX
DT 08-JUL-1996 (first entry)
DE C. difficile toxin B.
XX
KW Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin;
KW diarrhoea; therapy; diagnosis; vaccine.
XX
OS Clostridium difficile VPI strain 10463 (ATCC 10463).
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US13737.
XX
PR 07-JUN-1995; 95US-0480604.
PR 24-OCT-1994; 94US-0329154.
PR 16-MAR-1995; 95US-0405496.
PR 14-APR-1995; 95US-0422711.
XX (OPHI-) OPHIDIAN PHARM INC.
XX
PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
PI Williams JA;
XX
DR WPI; 1996-230603/23.

DR N-PSDB; AAT29247.
 XX Fusion proteins comprising non-toxin protein and part of toxin -
 PT useful to form anti-toxins against Clostridium botulinum type A, and
 PT C. difficile type toxins, and to treat C. difficile intoxication,
 PT partic. diarrhoea
 XX
 PS Claim 36; Page 313-323; 434pp; English.
 XX
 CC Clostridium difficile VPI strain 10463 toxin B (AAR95011), the
 CC product of the toxin B gene (AAT29247), is a cytotoxin associated
 CC with diarrhoeic disease. It can be obtd. by expression in
 CC transformed E. coli hosts of portions of DNA that together cover the
 CC entire toxin B gene. Toxin B, and portions of it (see also AAR95012-13,
 CC AAR95371-72 and AAR95018), pref. expressed as fusions to polystyridine
 CC affinity tags or maltose binding protein, are used to raise avian
 CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.
 XX
 SQ Sequence 2366 AA;

Query Match 77.6%; Score 6730.5; DB 17; Length 2366;
 Best Local Similarity 76.4%; Pred. No. 0;
 Matches 1298; Conservative 191; Mismatches 210; Indels 1; Gaps 1;

QY 1 MSLVNKAQQLQXVYVFRIOEDYVAILNALBEEYHNMSSSVVEKYKLKDINNLTDNYL 60
 DB 1 MSLVNRKQLEKMANVFRTOEDYVAILDALBEEYHNMSENTEVVEKYKLKDINSLTDYI 60

QY 61 NTYKSGRNKALKKFKXEYITMEVLELKNNSLTPVEKNLHFHWIGGQINDTAINYNQWKD 120
 DB 61 DTYKSGRNKALKKFKXEYITMEVLELKNNSLTPVEKNLHFHWIGGQINDTAINYNQWKD 120

QY 121 VNSDYTVKVFYDSNAFLINTLKTIVTESATNTLESFRENLDNDPEFYNKFKRMEIY 180
 DB 121 VNSDYNVNVFYDSNAFLINTLKTIVTESAINDTLESFRENLDNDPRFYNKFKRMEIY 180

QY 181 DKQKHFIDYKYSQIEENPEFIDINIIKTVLSNEYSKDLALNKYIEESLNKITTANNNDI 240
 DB 181 DKQKNFINYKAQREENPELLIDDIIVKTVLSNEYSKEIDELNTYIEESLNKITTQNSGNDV 240

QY 241 RNLEKPADEBLVRLYNQELVERWNLAASDIILRISMLKEDGGYVLDVDPGIDPLFKS 300
 DB 241 RNPEFKNGESFNLYEQELVERWNLAASDIILRISALKEIGWYLDVDMPLGIDPLFES 300

QY 301 INKPSDITWSEMIKLEALMKYKEYIPGYTSKNFDMLEEVQSPESALSKSKSEIF 360
 DB 301 IEKPSSTVTVDFWEMTKLEAMKYKEYIPEYTSHEFDMLEDEVQSPESVLAASKSEIF 360

QY 361 LPDDDIKVPLEVKIAFANNSVINOALISLKDSCYSDLVINOIKRYKILNDNLNPSINE 420
 DB 361 SSGDMEASPLEVKIAFNKGIINOGLISVKDSYCSNLIVKQIENRYKILNLSNLPASE 420

QY 421 GTDFNTMTKIFSKLASISNEONMMPMIKITNLYKVGFPADVRSTINLSGPGYTTGAYOD 480
 DB 421 DNDFTNTTTFIDSIMAEANADNGRPMELGKYLVRGFPDPVKTINLSGPEAYAAAYOD 480

QY 481 LLMFKNSTNIHLLEPELNFPEPKTKISOLTEQBITSLWSFNQARAKSQPEYKKGYPE 540
 DB 481 LLMFKEGSMNIHLIEADLNFNFESKTNISQSTEQENASLWSFDDARAKAQPEYKRNYPE 540

QY 541 GALGEDNDLDFACQTVLDKDYVSKKTLSSMKTRNKEYIHVIYVLOQGDKISYEASCNLFPSK 600
 DB 541 GSLGEDNDLDFSQNIIVDKEYELLEKISSLARSEGYIHVIYVLOQGDKISYEACNLFPSK 600

QY 601 DPVSSILYQKITEGSTAYVYVADAEIKEDIRIPYQISNKNRNLKFTFIGHGKSEFNT 660
 DB 601 TPYDSVLFOKNIEDSEIAYVYVNGDGEIQEIDKYKIPSIISDRPKIKLFTFIGHGKDEFNT 660

QY 661 DTFANLDVDSLSSEIETILNLAKADISPKYIEINLILGCMNMFYSIYAEETYPCKLLIKK 720
 DB 661 DIPAGFDVDSLSSEIAAIDLAKEDISPKSIEINLILGCMNMFYSINVEETYPCKLLIKK 720

QY 721 DRVSELMPSTSQDSITVSANQVEVRINEEGKREILDHSGKWINKEBSIIKDISSEYISF 780

DB 721 DKISELMPSTSQDSITVSANQVEVRINEEGKREILDHSGKWINKEBSIIKDISSEYISF 780
 QY 781 NPKENKIIVKSKYLHELSTLLOEIRNNANSDDIDLEKKVMLTECEINVASNIDRIVEGR 840
 DB 781 NPKENKIIVKSKYLHELSTLLOEIRNNANSDDIDLEKKVMLTECEINVASNIDRIVEGR 840
 QY 841 TEEAKNLTSDSINYIKNEFKLIESISLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
 DB 841 TEEAKNLTSDSINYIKNEFKLIESISLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
 QY 901 INKETGSGIFTEKEKEIFESEYATHISKEISNIKOTIFDNVNGKLVKKNLDAHVNTLN 960
 DB 901 INKETGESIFVETEXTIFESEYANHITEISIKIGIFDVGKLVKKNLDTTHEVNTLN 960
 QY 961 SAPPFQSLIEYNTTKESISNLSVAMKVQVYQALFSTGLNTITDASKVELVSTALDETID 1020
 DB 961 AAFPQSLIEYNTTKESISNLSVAMKVQVYQALFSTGLNTITDAAKVELVSTALDETID 1020

QY 1021 LLPTLSEGLPIATIIDGVSLGAAIKELSETNDPLLRQIEIAKIGIMAVNLTAATAVT 1080
 DB 1021 LLPTLSEGLPIATIIDGVSLGAAIKELSETSDPLRQIEIAKIGIMAVNLTTATTAIT 1080

QY 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETEGAFITLD 1140
 DB 1081 SSLGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVVDYFKHVSILVETEGVFTLTD 1140

QY 1141 DKIMPDQDLVLSIEDFNNSITLQKCEIWRAGSGSHLTDDIDHFFSSPSITVRKPMWL 1200
 DB 1141 DKIMPDQDLVLSIEDFNNSITLQKCEIWRMEGSGSHLTDDIDHFFSAPSITVREPHL 1200

QY 1201 STYDVLNITKKEIKDPSKDLMLPNAPRVFGVEMGTGPFSLDNDGFKLDIRDHVYG 1260
 DB 1201 STYDVLVLEQKELDUSKDLMLPNAPRVFAWETGTFELRSLDNDGFKLDIRDNDYEG 1260

QY 1261 QPYWRYFAFIADALITLKPREDYTNVRLNDGNTRSPFVPIITTEQIRKNLSYFSGG 1320
 DB 1261 EFWRYFAFIADALITLKPREDYTNVRLNDGNTRSPFVPIITTEYIREKLSYFSGG 1320

QY 1321 GYSLSLSPYNNIDLNVENDTVIDVNVVKNITIESDEIQKGLIENILSKUNIEDN 1380
 DB 1321 GTYALSLSQYNNGINIELSESDVMTIIDVNVVRDVTIESDKIKKGLIEGILSTLSIBEN 1380

QY 1381 KIILNHTINFVGDINESRFTSLTFSLEEDINIIIEIDLVSYSKILLSGNCKLIENS 1440
 DB 1381 KIILNSHEINFSGEVNGSGFVSLTFSLEGINAIEVDLLSKYSKILLSGELKILMLNS 1440

QY 1441 SDIOQKIDHIGFNGSHQKVIYFYSYIDNETKNGFIDYSKGLFTAEPNSESIIRNIYMP 1500
 DB 1441 NHIQKIDYIGFNSLOKNIPIYFVDSGKENGFGINGSTKEGLFVSELDPVVLIISKYMD 1500

QY 1501 DSNNFIYSSKDLKDIRINKGDVKLLIGNYFKDMKVSLSFTIEDTWTIKUNGVLDBN 1560
 DB 1501 DSKPSFGYVNNLKDVKITKDNVNILTYGVLKODIKISLTLQDEKTIKLSVHLDBN 1560

QY 1561 GVAQITLKPMNNAKSALNTSNLSMNPLESNTIKNFVNNLDNIEFILTDTNFIISGNSIG 1620
 DB 1561 GVAELTKPMNR-KGNTNTSDLSMNPLESNTIKSFVNNLQSNIKFILDANFIISGTSIG 1619

QY 1621 QPELTCDDKXIQPYFINFKIKETSYTLVGNRQNLIVEPSYHLDDSGNISSTVINFQK 1680
 DB 1620 QPEFICENDNIOQPYFIKNTLETNTYTLVGNRQNLIVEPYVLDLDDSGDISSTVINFQK 1679

QY 1681 YLYGIDRVNKKVIAPNLYT 1700
 DB 1680 YLYGIDSCVKNKVISPNLYT 1699

RESULT 2
 AAW68388
 ID AAW68388 standard; Protein; 2366 AA.
 XX
 AC AAW68388;

[illegible]

Db 1441 NHQQKIDYGFNSLQKNIPIYSFVDSKENGFGINGSTKEGLFVSELDVVLISKVYMD 1500
 Qy 1501 DSNLFYSSKOLKOTIRINKGVDKULLIGNYFKDMKVSLSFTIEDTNTIKLVGVYLDN 1560
 Db 1501 DSKPSFGYSSNNLKOVKITKONVNLTGTGYLKDDIKISLSLTLQDEKTIKLSNVHLDS 1560
 Qy 1561 GVAQILKFNNAKSALENTSLSMNFLESINIKNIFYNLDPNLEETLDTNFIISGSNSIG 1620
 Db 1561 GVAEILKFNMR-KGNTNTSDLSMFLSNMKSIFVNFLOSNIKFLDANFIISGTTISG 1619
 Qy 1621 QFELICDCKDNKTOPYFINKIKETSITLVGNRQNLIVERSYHLDDSGNISSTVINFSOK 1680
 Db 1620 QFELICDENDNTOPIYKENTLENTLVYGNRQNLIVERNVFNVDLDDSGDISSTVINFSOK 1679
 Qy 1681 YLYGIDRYNKVJIAPNLYT 1700
 Db 1680 YLYGIDSCVKNVISPNIYT 1699

RESULT 3

ID AAR95016 standard; Protein; 2710 AA.
 AC AAR95016;
 XX
 DT 08-JUL-1996 (first entry)
 XX
 DE C. difficile toxin A.
 XX
 KW Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;
 KW vaccine; diarrhoea; diagnosis; therapy.
 XX
 OS Clostridium difficile VPI strain 10463 (ATCC 10463).
 XX
 PN W09612802-A1.
 XX
 PD 02-MAY-1996.
 XX
 PF 23-OCT-1995; 95WO-US13737.
 XX
 PR 07-JUN-1995; 95US-0480604.
 PR 24-OCT-1994; 94US-0329154.
 PR 16-MAR-1995; 95US-0405496.
 PR 14-APR-1995; 95US-0422711.
 XX
 PA (OPHI-) OPHIDIAN PHARM INC.
 XX
 PI Pirca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
 PI Williams JA;
 XX
 DR WPI; 1996-230603/23.
 DR N-PSDB; AAT29248.
 XX
 PT Fusion proteins comprising non-toxin protein and part of toxin -
 PT useful to form anti-toxins against Clostridium botulinum type A, and
 PT C. difficile type toxins, and to treat C. difficile intoxication,
 PT partic. diarrhoea
 XX
 PS Claim 63; Page 290-302; 434pp; English.
 XX
 CC Clostridium difficile VPI strain 10463 toxin A (AAR95016), the
 CC product of the toxin A gene (AAT29248), is a potent cytotoxin that
 CC plays a direct role in damaging gastrointestinal tissues and is
 CC associated with diarrhoeic disease. It can be obtd. by expression in
 CC transformed E. coli hosts of portions of DNA that together cover the
 CC entire toxin A gene. Toxin A, and portions of it (see also
 CC AAR95014-15 and AAR95017), pref. expressed as fusions to polystyrene
 CC affinity tags or maltose binding protein, are used to raise avian
 CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.
 XX
 SQ Sequence 2710 AA;

Query Match 51.4%; Score 4458; DB 17; Length 2710;
 Best Local Similarity 50.7%; Pred. No. 1.6e-226;
 Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;
 Qy 1 MNLVNAQLOKQWVYKFRIOEDYVAILNALLEEYHNHSESSVVEYKYLKDOINLNTDYL 60
 Db 1 MSLISKEBELIKLAY-SIRPRENEYKTLTLDNEYNKLTNNENKYLQKKLNEIDVPM 59
 Qy 61 NTYKSGRGNKALKPKPEYLTMEVLELKNSTLTPVEKNLHFHWIGQINDTAINYNQWMD 120
 Db 60 NKYTSSNRNALSNLKDKILKEVILIKNSNTSPVEKNLHFHWIGGEVSDIALEYKQWAD 119
 Qy 121 VNSDYTVFVVDNAFLNTLTKTIVESATNTNLTLESFRENLDNDFEDYKFKYKMEIITY 180
 Db 120 INAEYNIKLWDSAEFLVTLTKAIVESSTTEALQLLEEBEQNPQDNMKFYKKMEFIY 179
 Qy 181 DRQKHFDIYKSIQIBENPEFIIDNIITKYLSEYKSKOLEALNKYTEESLNKITTANNNDI 240
 Db 180 DRQKEFINYKSIQINKPTVIDDIKSHLVSEYNRDETVELSYRTNLSRKINSNHGDI 239
 Qy 241 RNLEKFADEDLVRLYNQBELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDPLPKS 300
 Db 240 RANSLFTEQELLNYSQELNARGLAASDVRLLALKNFGVYLDVDMPLPGIHSDFKT 299
 Qy 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSSKSKSEIF 360
 Db 300 ISRPSSIGLDREWEMIKLEAIMKYKEYINNYTSENFDKDDQQLQKFKLIIESKSEIF 359
 Qy 361 LPDDIKVPLEVKIAFANNVINOALISLKDSCYSDLVINQIKNRYKILNDNLNPSINE 420
 Db 360 SKLENLVSDLEIKIAPALGVSINOALISKQGSYLTNLVIEQVKRYQFLNQHLPALIES 419
 Qy 421 GTDFNTWKIFSDKLASISNEDNMFMKIKNYLVKGFAPDVRSTINISGPGVYTGAYQD 480
 Db 420 DNNFTDTTKIFHDSLFNSATAENSMTKIAPIYQVGFMPPEARSTISLSPGAYASAYD 479
 Qy 481 LLMFKDNSTNIHLLEPELNFEEFKTKISQTEQITSLMSNOARAKOSEEYKGYFE 540
 Db 480 FINQENTIEKTLKASDLIEFKPPENNLSLQTEQEINLSWSPDOASAKYQEKYVRDVTG 539
 Qy 541 GALGEDNDLPAQNTVLDDKY-VSKILSS--MKTRNKEYIHYIYVLOQDKISYEASCNL 597
 Db 540 GSLSEDNGVDNFKNATLDKNVLLNKNKIPSNVVEEAGSKNYHYIYVLOQDDISYEATCNL 599
 Qy 598 FSKDPYSILYQKNIEGSETAYYYVAD--AEIKEDIKYRIPYQISKNKIKLTFIGHGK 655
 Db 600 FSKNPKNIIQIRNM--NESAKSYFLSDDGESILLEKNKYRIPERLKNKEKVKVTFIGHGK 657
 Qy 656 SEFNTDTFANLDVDSLSSEIETILNLAADISPKYIEINLGCNMFSYIYAEETYPGKL 715
 Db 658 DEFNTSEFARLSVDSLSNEISFELDTIKLDISPKNVEVNLGCNMFSYDFNVEETYPGKL 717
 Qy 716 LLKIKDRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWKINKEESIIDKSSK 775
 Db 718 LLSIMDKITSLPDVKNKSITIGANQYEVRIINSEGRKELLAHSGKWKINKEEAIMSDLSK 777
 Qy 776 EYISNPENKILYVSKYLHELSTLQBIIRNANSDDILEKQVMLTCEINVASNIDRQ 835
 Db 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLDDASVSPDTKTLNKLNIIESIGDY 837
 Qy 836 IVEGRIEBAKNTSDSINYIKNEFKLIESISLSYDLAKHQGLDDSHFISFEDISKTEG 895
 Db 838 IYKEKLEVPKNIHNSIDDLIDEFNLENNVSDLEYELKUNLNDKYLISFEDISKNNST 897
 Qy 896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNIMKDTIFDNVNGKLVKKVNLDAHAH 955
 Db 898 YSVRFINKSGESVYVEKEIFSKYSEHITKEISTIKNSITDVGNNLDNIQLDHTSQ 957
 Qy 956 VNTLNSAFIQSLIEYNTTKESLSNLSVAMQVYQAQLFSTGLNTITDASKVVELVSTAL 1015
 Db 958 VNTLNAAPFIQSLIDYSSNKVDLNTSVKQVLAQLFSTGLNTIYDSIQLVNLISNAV 1017
 Qy 1016 DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETNPDLLRQRETEAKIGIMAVNLTAAS 1075

600 FSKNPKNSIIQRM--NESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGK 657
656 SERNTDTFANLDVDSLSSEIETILNLAKADISPKYIEINLGCNMFPSYIYAEETYPGKL 715
658 DEFNTSEFARLSVDSLSNBISSFLDKIDISPKNVEVNLGCNMFPSYDFNVEETYPGKL 717
716 LKIKDRVSELMPISQDSITVSANQYEVINEEGKREILDSHGKWINKEESIIKDISK 775
718 LLSIMDKITSTLPDVKNKISITIGANQYEVINESEGRKELLASHGKWINKEEAIMSDLSK 777
776 EYISFNPKNKIIVGSKYLHEISSLQEIERNNANSSDIDLEKKVMLTECEINVASNIDRQ 835
778 EYIFPSIDNKLKAKSNIPGLASISEDIKTLDDASVSPDKFILNKLNIIESIGDY 837
836 IYEGRIEEAKNLTSDSINIKNEFKLIESISDSLYDLKHQGLDSDSHPISEFIDISKENG 895
838 IYEEKLEPKNIHNSIDILDEFNLLENVSDLEYELKKNLDEKYLISFEDISKNST 897
896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAHE 955
898 YSVRFINKSNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLNLDNIQLDHTSQ 957
956 VNTLNSAFFIQSLIENYTTKESLNSLVAMQVYAQLFSTGLNTITDASKVVELVSTAL 1015
958 VNTLNAAFFIQSLIDYSSNKVDNLDSVSKVQVLAQLFSTGLNTIYDSIQLVNLISNAV 1017
1016 DETIDLPLTSEGLPIATIIDVSLGAAIKELSETNDPLLRQETEAIGIMAVNLTAAS 1075
1018 NDTINVLPTITGIPVSTILGINLGAALKELLDEHPLLKELEAKVGLVAINMSLSI 1077
1076 TAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILODKATVIDPKHISLAETEGA 1135
1078 AATVASIVGIGAEVITFLPIAGISAGIPSLVNNELIHDKATSVVYFNHLSSEKGYG 1137
1136 FTLLDDKIIMPODVLSEIDFNNSITLCKEINRAEGSGHTLTDIDHFFSPSITY 1195
1138 LKTEDDKILVPIDDLVISIDFNNSIKLGTGNILAMEGSGHTVTGNIDHFFSPSIS 1197
1196 RKPWLSIYDLNLIKKEIDFSDKDLMLVLPNAPRVFGEYENGWTPGFRSLNDGDKLLDRR 1255
1198 HIPSUSIYSAIGIETENLDFSKIMMLPNAPRVFWETGAVPGURSLNDGTRLLDSIR 1257
1256 DHYEGOFYWRFAFIADALITKLKPRYEDTNVRINLDGTRSFIVPVIITEQIRKNLSYS 1315
1258 DLYPGKFYRFPYAFF-DYAITLTKPYVEDTNIKIKLDKTRFIMPTITNIRKNLSYS 1316
1316 FYVSGGSYSLSPYNNIDNLVENDTVIDVNVKNITIESDBIQGELIENILSKL 1375
1317 FDCAGGYSILLSSYPISNTINLSKDDLWIFNIDNEVREISIENTGTIKGKLKIDVLSKI 1376
1376 NIEDNKILNNHTINFYGDINESNREISLTFSLEIDINIIIEIDLVSYSKYILLSGNCWK 1435
1377 DINKNKLIIIGNQTDIFSGDIDNDKORYIFUTCELDLDDKISLIIIEINLVAKSYSLLSGDKNY 1436
1436 LIENSSDIOQKHIDHGFNGEHQKIFVSYVID-NETKNGFIDYSKKEGLFTAEFNSNESTII 1494
1437 LISLNSNTEKINTGLD---SKNIAVNTDSNKKYFGAI-----SKTSQKSI 1483
1495 RNITMPDSNNL-----FIYSSKDL--KDRIINKGDVKKLIGNIFYKD---DMKVSUS 1541
1484 H--YKDSKNILFEYNDSTLEFNSKDFIAEDINVFMKDDINTITGKYVYDNNNTDKSIDFS 1541
1542 FTIEDTNTIKLGVYLDENGVAQILKFMNNAKSALNTSNLNFLESINIKIIFYNNLDP 1601
1542 ISLVSNQVKNVGLYNESVYSYLDVFNKSDGHNTSNFMNLFNINISFWLFGPE--- 1598
1602 NTEFIDLTNFIISGNSIGOFELICDKNIOPIYFNFKIKTSYTYLVGNRQNLIVEPS 1661
1599 NINFVIDKYFTLVGKTNLGVVEFICDNNKNIDIFYGCEWTKSSKSTIFSGNGRNVVVEPI 1658
1662 YHLDDSGNSTSVINFSQKLYGIDRYVANKVIAPNLYT 1700
1659 YN-PDTGEDISTSLDPSYPLYGIDRYINKVLIAPDLYT 1696

RESULT 5

AAV33700

ID AAV33700 standard; protein; 546 AA.

XX

AC AAV33700;

XX

DT 18-JAN-2000 (first entry)

XX

DE C. sordellii lethal toxin protein fragment.

XX

KW Lethal toxin; immunotoxin; antitumor; Glucosyl transferase; glycosylate;

KW GTPase; Ras; epidermal growth factor stimulated MAP-kinase; inhibition;

KW signalling pathway; cell-specific toxin; treatment; cancer.

XX

OS Clostridium sordellii.

XX

PN DE19802569-A1.

XX

PD 09-SEP-1999.

XX

PF 23-JAN-1998; 98DE-1002569.

XX

PR 23-JAN-1998; 98DE-1002569.

XX

PA (UYFR-) UNIV FREIBURG ALBERT-LUDWIGS.

PI

Aktories K, Hofmann F;

XX

DR WPI: 1999-509323/43.

XX

N-PSDB; AA223800.

XX

PT New fragment of the lethal toxin from Clostridium bacterium, useful for

XX

treating cancer -

XX

PS Claim 1; Page 7-9; 14pp; German.

XX

This invention describes a novel fragment (I) of the lethal toxin (LT) of Clostridium sordellii which has antitumor activity. (I) is a glucosyl transferase that glycosylates, and thus inactivates, GTP(guanine triphosphate)ases, particularly Ras (an oncogenic product overexpressed in many tumors), resulting in inhibition of epidermal growth factor stimulated MAP-kinase signalling pathways. (I), particularly in the form of immunotoxins, are used as cell-specific toxins, particularly for treating cancer. When included in immunotoxins, (I) can be targeted to selected cells. Compared with the complete LT, (I) is smaller, so enters cells more easily, resulting in greater toxicity in the cytosol, is less likely to induce formation of (neutralizing) antibodies, and is more active than the holotoxin. This sequence represents the lethal toxin fragment described in the invention.

XX

SQ Sequence 546 AA;

XX

Query Match 32.0%; Score 2780; DB 20; Length 546;

Best Local Similarity 99.1%; Pred No. 8.2e-139;

Matches 541; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY

1 MNLVNAQLQKVVYKFRIOEDYVAILNLAEEYHNMSSESVVEKYLKLDINNTDNYL 60

DB

1 MNLVNAQLQKVVYKFRIOEDYVAILNLAEEYHNMSSESVVEKYLKLDINNTDNYL 60

QY

61 NTKKSGNKKALKKPKVEYLTWVEVLKNNSLTPVEKNLHFIIWGGQINDTAINYNQWKD 120

DB

61 NTKKSGNKKALKKPKVEYLTWVEVLKNNSLTPVEKNLHFIIWGGQINDTAINYNQWKD 120

QY

121 VNSDYTVKVFVDSNAPLNTLKTIVESATNTNLTSPRENLDPEFDYKFKRKEIYY 180

DB

121 VNSDYTVKVFVDSNAPLNTLKTIVESATNTNLTSPRENLDPEFDYKFKRKEIYY 180

QY

181 DKQKHFDIYKQSIEENPEFIIDNIKTLYSGNEYSKOLEALNKYTEESLNKITTANNNDI 240

DB

181 DKQKHFDIYKQSIEENPEFIIDNIKTLYSGNEYSKOLEALNKYTEESLNKITTANNNDI 240

QY 241 RNLEKFAEDLVRLNQELVERWNLAAASDIIRISMLKEDGGVYLDVILPGIQDPLFKS 300
 DB 241 RNLEKFAEDLVRLNQELVERWNLAAASDIIRISMLKEDGGVYLDVILPGIQDPLFKS 300
 QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEDEEVORFESALSSKSKSEIF 360
 DB 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEDEEVORFESALSSKSKSEIF 360
 QY 361 LPDDIKVSPLEVKIAFANNSVINQALISLKDSCYSDLVINOIKNRYKILNDNLNPSINE 420
 DB 361 LPDDIKVSPLEVKIAFANNSVINQALISLKDSCYSDLVINOIKNRYKILNDNLNPSINE 420
 QY 421 GTDFNTTMMKIFSDKLASINEDNMFMKIKITNYLKVGFAPDVRSTINLSGPGVYTGAYQD 480
 DB 421 GTDFNTTMMKIFSDKLASINEDNMFMKIKITNYLKVGFAPDVRSTINLSGPGVYTGAYQD 480
 QY 481 LLMFKDNTNIIHLLPELNFPPKTKISQLEBISLWSFNQARAKSQFEYKKGYPF 540
 DB 481 LLMFKDNTNIIHLLPELNFPPKTKISQLEBISLWSFNQARAKSQFEYKKGYPF 540
 QY 541 GALGED 546
 DB 541 GALGED 546

RESULT 6

AA18171
 ID AA18171 standard; Protein; 1979 AA.
 XX
 AC AA18171;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:28.
 XX
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200025728-A2.
 XX
 PD 11-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26796.
 XX
 PR 05-NOV-1998; 98US-0107131.
 XX
 PA (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX
 DR WPI; 2000-365347/31.
 XX
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection -
 XX
 PS Disclosure; Page 70-75; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)
 CC vaccines against P. falciparum infection comprising (i) or (ii).
 CC (i) and (ii) are useful for the development of vaccines against
 CC P. falciparum infection. (i) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (i), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (i) (especially when they are rifins or secreted or membrane proteins)

CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the plasmidium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasitic life cycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.
 XX
 SQ Sequence 1979 AA;
 Query Match 4.6%; Score 398; DB 21; Length 1979;
 Best Local Similarity 19.7%; Pred. No. 2.9e-12;
 Matches 373; Conservative 287; Mismatches 605; Indels 626; Gaps 88;
 QY 53 NNLTNYLNTYKSGRNKALKKFKYELTMEVLELKNLSLTPVEKNLHPTWIGQINDTAI 112
 DB 93 NNNDNN 149
 QY 113 NYINQWQVNSDYTVKPYDSNA---FLINTLKTIVESAATNTLTSPREN---LNDPEF 166
 DB 150 NYESKIEELEKE--LKEVKDKNIIDNNYENKLEK--EDFVKQKIDMLNEKENLQEKEL 205
 QY 167 DYNKPYRK---RMEIYDKQKHFDYKSOEENPE-----FIIDNIKTLYS----- 211
 DB 206 DINKREKINKENKNIKKSETFHNIKEYLEKNERETISIEIID--IKGHELEKLIBI 263
 QY 212 NEYSKDLEALK------YTESLNKITTANNNDIRNLEKFADELDVRLYNO 257
 DB 264 KEKEDLENLKNLLSKENLVKELKGCVKERNETINSLNDNIEKKY-----KLLEY 317
 QY 258 ELVERWNLAAASDIIRISMLKEDGGVYLDVILPGIQDPLFKSINKPSITNTSWEMIKL 317
 DB 318 ELEEK-----NKQIDL---LNKQEKKEKEKEKEKE 345
 QY 318 EAIMKYKEYIPGYTSKNFDMLEDEEVORFESALSSKSKSEIFLPLDDIKVSPLEVKIAP 377
 DB 346 KEREKEKE-----KEYDTLTKEL-----KDEKISILEKVHISIKVREMDIERKE 388
 QY 378 ANNSVINQALISLKDSCYSDLVINOIKNRYKILNDNLNPSINEGTDFTNTMKIFSDKLAS 437
 DB 389 HNFLHMEQDLKDLKNSFVN--NNQLK-VYKCEIKNLK-----TELEKEKELKD--TEN 438
 QY 438 ISNEDNMFMKIKITNYLKVGFAPDVRSTINLSGPGVYTGAYODLLMFKDNTNIIHLLPE 497
 DB 439 VSKEE---INKLINQLN-----EKEKQILAFNKNHKEIHGLKEE 475
 QY 498 LNFEPFPKTKISQLEBISLWSFNQ-----ARAKSQFEYK 536
 DB 476 LKE-----SVKTKTETQLEQEMVDIKQELDQEKYNAQIESISIELSKKEKYNQYN 531
 QY 537 GYFEGALGEDNDLDPQNTVLDKDVSKILLSSMKTRNKEYIHYIVQLQGDKISYEASCN 596
 DB 532 TYIE---EINNLN-----EKLEETNKEY---TNLQN---NTNEIN 563
 QY 597 LFSKDPYSGIYOKNIEGSETAYYYYVADAEIKEDYRIPYQISNKRNIKLTFIGHGKS 656
 DB 564 MLNND-----IHLNGLNKTWNT-----QISTLKN----- 588
 QY 657 BNTDTFANLDVDSLSSEIETILNLAADIPKYTEINLGLCNMFSYIYABETVPGKLL 716
 DB 589 -----DVHLLNEQIDK-LNNEKGTLSKISELVN-----QI 618
 QY 717 LKIKDRVSELMPISQDSITVYSANQYEV--RINEGKEILDHSGKWINKEESIUKDSS 774
 DB 619 MDLKEEKDFL-----NNQIVLDSNQIDLLTRQEEKENKMLEQENKYQEMELLRGNIKS 673
 QY 775 KEYISFNPENKNIIVKSKYLHELSTLQEIIRNNANSSOIDLEKKVMLTECEIN--VASNID 833

Db 674 SENILNDBE-----VCDLKRKLSLSESEMKMKKEHD 706
Qy 834 ROIVE-----GRIEAKNLTSDSINYIKNEFKLIESISDSLYDLKHONGLDDSHFISPE 887
Db 707 KLAELKDDCVIRIEMNEKEDKINMLKEEY-----DKINTLKEQN-----E 750
Qy 888 DISKTENGFRIRIPINKETGNSPIETEKEIFSEYATHISKEISNIKDTIFDNVNGKLKVK 947
Db 751 D-----KINTLKEQ--NEDKINTLKE--EY-----BHKINTMKEEYEHEKIN-----T 788
Qy 948 VNLDAAEVNTLNSAFFIQLSLEYNTKESLNSLVAM-----KVQVVAQLFSTGLNTI 1001
Db 789 LNEQNEHKINTLINE-----QNEHKINTMKEEYEDKMTLNEQNEKDNLSKEEYENKINOI 844
Qy 1002 TDAS--KVVELASTALDETIDLPITLSEGLPIIATIIDGVSILGAALKSELSETNDPLRQE 1059
Db 845 NNNEIKIKDVVNEYIEEVKLVTLDEKKKQFDKEINYAHKAHEKE-----892
Qy 1060 IEAKIGIMAVNLTAATAIIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILODKATK 1119
Db 893 -----QILLTEM-----EELKCO-RDNK 909
Qy 1120 VID-YPKHISLAETAGFTLDDKIIMPQDDVLVLSIDFNNSITLTKCEIWRAB--GGS 1176
Db 910 YSDLYEKYIKLKSICMIINIECCDDIENEDIIRIEEYINNKKGLK-EVEKEKHKRS 968
Qy 1177 GHTLTDDIDHFFSSPSITRKWLSIYD---VLNIKKEKIDFSLKOLWLPNAPNRFVGE 1233
Db 969 SPNLSKKEKFPN-----SIEDKSHLKKKHEKDLSSKKEI--EEKNK-----1011
Qy 1234 MGWTFGRSLDNDGTGLLDRIR-----DHVEGQFYWRYFAFIADALITLKLKPRY 1282
Db 1012 -----KIKELANDIKKLQDEILLVYKQSNAAQVDHKKKSWI-----LLKDKSKKEI 1057
Qy 1283 EDTNVRINDGNTRSPI-----VPVITTEQIR-KWLSYSFYGSGSYLSLSPYNNMID 1335
Db 1058 KOKENQINVEKEEKDLKKDDKDEIRILNEELVKYTI-----LYNLKKDPLQON 1108
Qy 1336 -LNLVENDTWID---VDNVKNITIESEDEKQK-----ELIENILS-----KLNTED 1379
Db 1109 LUSKDINDSLTNEGCMVDKIEHILDYDEEINKSRNLFQKNEICSTTTEWELNNKK 1168
Qy 1380 NKIILNHTINFG-----DINESNRFISLTPSILEDINIIE-----I 1418
Db 1169 NELIEENKLNLDVQGGKKLKDVEKQKKEIEKLNKQLTKCNQIDELNEVEKLNENI 1228
Qy 1419 DLVSKS-----YKILLSGNCKLIENSDDIOQ---KIDH-----1449
Db 1229 ELITYSNDLNNKPFMKENLNMKLDENEDNIKMKSKIDDMKEIKYREDEKKRNLINEIN 1288
Qy 1450 -----IGFNGEHQKY-----IFYSYIDNETKYNFP-IDYSKK-----1480
Db 1289 NLKKKNEDWCIKYNEKNIKYDGVCKYBEMSLTYKETSLSKYBQIKVYDEKCSQVDENF 1348
Qy 1481 ---EGLFTAFFSNE---SIIRNIYMPDSNNLFTYSSKDLKDRIINKGDVKKLIGNYFK-- 1533
Db 1349 QYDEKCFQDEINKKYGALLNI---NITNKMVDSKVDRNNNELLISVDNKKVEGIANYLKOI 1405
Qy 1534 -----DDMKVLSLFT---IBDTNTIKLNGVYLDENGVAQILKFM-----NNAKS 1574
Db 1406 FELNEEIRLKGBEINKISLLYSNELNEKNSYDINMKHIOEQ-----LLPLEKTNKENERK 1460
Qy 1575 ALNTNSLNM-----FLESINI-KNIFYNLDPN---IFFILDTNFIISG 1615
Db 1461 IINLTSQYSDAYKKKSDESKLCAQFVDVNIYGNISNNIRTNNEYKYBEMFDTNIEKN 1520
Qy 1616 SNSIGOFELICDKK--NIPQYFINKIKET-----SYTLVGNRQNLIVE---PSYH 1663
Db 1521 GHLSKYIHLLENFRCKMIYENENIKSSNKIIGLYNSRYGLREDLCKEEIVPS---1578
Qy 1664 LDGSGNISSTVINFOSKYLYGIDRYNVKVII 1694
Db 1579 --KIGNISKNENNNKKN-NTCDGYDEKVTI 1606

RESULT 7

AAAG82935
XX AAG82935 standard; Protein; 5024 AA.
XX AAG82935;
XX AC
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:2964.
XX KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
XX KW vaccination; endocarditis.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US30782.
XX PR 09-NOV-1999; 99US-0164258.
XX PA (GLAX) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX DR WPI; 2001-316495/33.
XX N-PSDB; AAH53785.
XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX Claim 18; Page 779-781; 2188pp; English.
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AG81454 to AG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX SQ Sequence 5024 AA;
Query Match 4.1%; Score 357.5; DB 22; Length 5024;
Best Local Similarity 19.5%; Pred No. 1.4e-09;
Matches 367; Conservative 310; Mismatches 631; Indels 573; Gaps 100;
Qy 20 QDEYVAILNALAEYHNHNSSESVVEKYLKLDINN-----LTDN-----YLN-- 61
Db 2611 OKD---TILNHI--PSAPTRSOVGKIASAKOLNNTMKALRDSIADNNILQSSKYFNEB 2665
Qy 62 TYKSGRNKALKKFKYELTMEVLELKNLSLPV-----EKNLHFIIWIGG 105
Db 2666 SEQQNAVQAVNKAKNII-----NQDPTVMADEIQSVLNEVKOTKDLH-----GD 2713
Qy 106 Q--IND-----TAINYINQWKNVSDYTVKVFYDSN-----AFLNLTCKT 144

Db 2714 QKLANDKTDQAATLALNALYNOAQRGNLETKVQ---NSNSRPEVQKVQVLANQLANDAMKK 2770
 Qy 145 IVESATNTLESPRENLDNPEFYKFKRMEIILYDKQKHFDYKQIIE--NPEF-- 200
 Db 2771 LDALUTGN--DAIKQTSYINEDTSQ-----QVNPDE---YDRGKNIVAEQTPNPNMSP 2819
 Qy 201 -----IIDNIITKYSNYSKDEALNKYIEBSLNKITYANNNDIRNLEKFADEDLVRLY 255
 Db 2820 TWINTIADKI-----TEAKNDLHGQVKL--EQAQQQSINTINQMTGLNQAQKEQL---- 2867
 Qy 256 NOELVERMNLAAASDILRISMLKBDGGVYVDVILPGIQDPDFKSINKPDSITNTSWEMI 315
 Db 2868 NOEIQOTQTRSEVHVINKAALNDS-----MNTLRQSITDEHEVKQTSNYIN 2915
 Qy 316 KL-----EAIMKYEVI PGYTSKNFMDLDEBQVRSFESALSSKSKSEIFLPLDDIK 367
 Db 2916 ETVGQATYNNADVVKQIINOTSNTMPL--EVERATSNVKTSK-DALHGERELNDNK 2972
 Qy 368 VSPLEVKIAPANNV--INQALISLKSYSYCDL-----VINQIKNRYKIILN----- 411
 Db 2973 NSK-----TFVNLHLDNLQA--QKEALTHEIQATIVSQVNNIYNKAKALNNDMKKLK 3024
 Qy 412 -----DNLNPSINEGTDFNTWKIPSPDKLASISNEDNNMFI-----KITNYLVKG 457
 Db 3025 DIVAQODNVRSNNYINEDSTPQMYND--TINHAQSIIDQVANTPTMSHDEIENAIN-- 3079
 Qy 458 PAPDVRSTIN-LSGPGVYGAQDMLFMKDNSTNHLPELNPPEFP-KTKISOLTE-- 513
 Db 3080 ---NIKHAINALDGE-----HKLQQAQKANA---NLLNSLNDLNAPORDALINLNEA 3126
 Qy 514 -----QBITSLMSFNQA-----RAKSQF---EYVKGYFEGALGEDNL 549
 Db 3127 QTRKEVAEQLOQAQALNDAMKHLRNSIQNSSVRQESKYINASDAKKEQYNHAREVENI 3186
 Qy 550 DPAQNTVLQDYVSKILSSMKTRNKEYIHYIVLQGDKISYEASC----- 595
 Db 3187 INEQHTPLDKETI-KQLTDAVNOAND-LNGVELLDADKQNAHQSIPTLMHLNQAQNAL 3244
 Qy 596 -----NLFSDKDPYSILYQ-----KNIESEATYIYVADAE-IKEIDKVRPYQIS 641
 Db 3245 NEKINNNAVTRAKVAALIGQAKILDHAMEDNES-----IKDKEQVQSSNY-----IN 3292
 Qy 642 NKRNIKLF---IGHCKSEFNTDFTANLDVDSLSSEITILMLAKADISPKYIEINLLGC 698
 Db 3293 EPDQVQETTNNAVHVEITLQNTVNTPLSIEDIEHAINEV-NOAKKQLRGK----- 3342
 Qy 699 NMFSYSIYAETYPGKLLIKIDRVSELMPSISQDSITVSANQYEVINEEGKR----- 752
 Db 3343 ----QKLYQTDIDLADKELSKLDLTSQSSSISNQIYTTAKTRTEVAQAIEKAKSLNHAMK 3398
 Qy 753 -----EILDHSGKWINK-----ESTIKDISKEYISFPNKENKILVSKYLH 795
 Db 3399 ALNKIYKNADKVLDD-SRRFINEDQPEKAYQQAINHVDVSIHRQTNPEMDPTVINS-ITH 3456
 Qy 796 ELSTLLQEIERN-----ANSSDIDLEKKVMLTECEIN-----VASNIDR- 834
 Db 3457 ELETAGNNLHGDKLAHAKQDAANVINGLIHLNVAQREVMINTNTATREKVAKULDNA 3516
 Qy 835 QIVEGRIEAKNLTSIDINYNKPEKLIBESI SDSLVDLKHONGLDSDHFSIFEDISKTEN 894
 Db 3517 QALDKAMETLQOVVAHK-NNILLNDSKYLN--EDSKYQQQVDRVIADAEOQLLQNTTPTLE 3573
 Qy 895 GFPIRFINKETGNSIETIEKEIFSEYATHISKEISNIKDTIPDNVNGKLVKNVNLDAH 954
 Db 3574 PYKVDIVKDNV-----LANEKILFG--AEKLSYDKSNANDEI-----K 3609
 Qy 955 EVNTLNSAFFIQSLIEYNTTKBSLNSLVAMKVQVVAQLFST---GLNTITDASKVVELV 1011
 Db 3610 HNNYLNNA-----OKQSIKIMISHAALRTEVKQLLOQAKTLDKAMKSLDKTQVV-IT 3661
 Qy 1012 STAL-----DETIDLLPTLSEGLPIIATII--DGVSLG---AAIKELSETNDPL-- 1055

Db 3662 DTTLPNRYTEASEDKKEKVD--QTVSHAQAIIDKINGSNVSLDQVRQALEQLTOAENLDG 3719
 Qy 1056 -LRQIEIAKI-----GIMAVNLTAATA--IVTSALGIASGFSILLVPLAGISAGIPS 1105
 Db 3720 DQREVE-EAKVHANQTIIDOLTHLNSLQQQTAKESVKNATKLB-----TAT 3763
 Qy 1106 LVNNELIQQDKATKVIDYFKHISLAETEGAFLLDDKIIMPDDLVLSLSEIDPNNNSITIG 1165
 Db 3764 ASNNALANLKVWGKLEQFINHADSIENSNDYRQADDKIIAYDDALEHGQDIQKSNATQN 3823
 Qy 1166 KCEIWRAGSGSGHTLTDDIDHP-FSSP-SITYRKFWLSIYDVNLNKK-----EKIDFSKDL 1219
 Db 3824 EAKQALQULINAEATSLNGFERLNHARPRALEYIK---SLEKINNAAKSALEDKVTQSHDL 3880
 Qy 1220 MVLNAPNRVFCYENGWTPGFRSLNDGCTKLDRIRDHVEGQFYWRYFAFTADALITKLK 1279
 Db 3881 LELEH-----LVNEGTLNDINGE-----LANAIVN--- 3906
 Qy 1280 PRYEDTNVRIINLDGNTRSFIVPITTEQIRK-NLSYSGYSGYSLSLSPYNNMIDNML 1338
 Db 3907 -NYAPTASIN-----YINADNLKONFTQAINNARDALN---KTQCNLDENA 3951
 Qy 1339 VENOTWVIDVD-----NVVKNIITIEDEIOKGLIENI--LSKLNI-EDNKIILNNHTI 1389
 Db 3952 I--DTFKDDIFKTKDALNGIERLTAAKAKAEK--LIDSLKFINKAQFTHANDEIMWNTNS- 4006
 Qy 1390 NPYGDIENSRFISLTFISLEIDINIIEIDLVSYSKILLSGCMKLIENSDDIOQKIDH 1449
 Db 4007 -----IAQLSRIVNQAFD-LNDAMKSLRDELNNQAPFVQASN---YINSDEDLKQQDPH 4057
 Qy 1450 IGFNGEHQKIFYSYIDNETKYNFYDKYKGLFTAFESNIESIIRNIYMPDSNNLFIYS 1509
 Db 4058 A-----LSNARKV-----LAKENG----- 4071
 Qy 1510 SKDLKDRIIINKGDKVLLIGNYFKDDMKVLSFTIEDTNTIKLNGVYLDNGVA----- 1563
 Db 4072 -KNLDEIQI-----EGLQVIEDTKD-ALNGIQRLSKAKAKAIQYV 4110
 Qy 1564 QILKEMNNAKSALNTSN-----SLMNFLESINIKNIFYNNLDNPTIEFIDTNTFIISG 1615
 Db 4111 QSLSYINDAQRIHIAESNIHNSDDLSLANTLSKA-----SDLDNAMKDLRDT--LESN 4161
 Qy 1616 SNSIQGFELICDKOKNIQPYF 1636
 Db 4162 STSVSPSVNYINADKNLQIEP 4182
 RESULT 8
 ABP38314
 ID ABP38314 standard; Protein; 10182 AA.
 XX
 AC ABP38314;
 XX
 XX 24-JUL-2002 (first entry)
 XX
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3159.
 XX
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy.
 XX
 OS Staphylococcus epidermidis.
 XX
 XX US6380370-B1.
 XX
 PD 30-APR-2002.
 XX
 PF 13-AUG-1998; 98US-0134001.
 XX
 PR 14-AUG-1997; 97US-055779P.
 PR 08-NOV-1997; 97US-064964P.
 XX
 XX (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;
XX WPI; 2002-381255/41.
DR N-PSDB; ABN90859.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 3159; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 10182 AA;

Query Match 4.0%; Score 348; DB 23; Length 10182;
Best Local Similarity 19.4%; Pred. No. 1.1e-08;
Matches 354; Conservative 317; Mismatches 663; Indels 494; Gaps 97;

QY 20 QDEYVAALLNALLEYHNMSESSVVEKVLKLDINN-----LTQN-----YLN-- 61
DB 7557 QKD---TILNHI--FSAPTRSQVGEKIASAKQLNNTMKALRSDIADNNEILOSKSYFNE 7611
QY 62 TYKSGRNRKALKKFEYLWVLELKNLSLTPV-----EKNLHFTWIGG 105
DB 7612 SEQQNAYNQAVKAKII-----NDQPTPWANDEIQSVLNEVKVQTRKDLNH-----GD 7659
QY 106 Q--IND-----TAINYINQWKNVSDYTVKFVYDSN-----AFLINPKKT 144
DB 7660 QKLANDKTDQAATNALNVLNQAGRNLETKVQ---NSNSRPEVQVQVQLANQLNDAMCK 7716
QY 145 IVESATNNTLESFRENLDPEFYNKFKYKRMIIIVDKQKHFDYKQIEE--NPEF-- 200
DB 7717 LDDALTGN--DAIKQTSYINEDTSQ-----QVNFDE---YTDGRKNIVASQTPNMS 7765
QY 201 ----TIDNIIKTYLSNEYSKOLEALNKY---IEESLNKITANNQNDIRNLEKFADEDIV 252
DB 7766 TWINTIADKI-----TEAKNDLHGQVKLQKQAOQQQSINTI-----NQMTGLNQAOKEQL- 7813
QY 253 RLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPFIQPOLFKSINKPDSITNTSW 312
DB 7814 ---NQBIQQTQTRSEVHVQVINKAALNDS-----MNTLRQSIITDEHVQKQTSN 7858
QY 313 EMIKL-----EAMTKYEIPGVTSKNFMDLDEEVQSRFESALSKSKSEIFLPJD 364
DB 7859 YINETVGNQTAYNNAVDRVKQIINQTSNPTMPL--EVERATSNVKISK--DALHGERELN 7915
QY 365 DIKVSPLVKIAPANNSV--INQALISLSDSYCSDL-----VINQIKNRVKYLN---- 411
DB 7916 DNKSK-----TFAVNHLNLAQA---QREALTHEIQATVISOVNNIYKAKALNNDMK 7967
QY 412 -----DNLNPSINEGTFNTMTKIFSDKLASINEDNMFMFI-----KITNYL 454
DB 7968 KKLKDIVAQQDNVRQSNYINEDSTPONMYND--TINHAQSIIDQVANPTMSHDEIENAI 8024
QY 455 KYGFAPDVRSTIN--LSCPGVYTGAYODLLMFQDNSTNIHLPELNFEPF--KTKISQIT 512
DB 8025 N-----NIKHAINALDGE-----HKLQQAQENA---NLLINSNLNAPQORDAIRNLV 8069
QY 513 E-----QBITLSWSPNQA-----RAKSRF---EEYKKGVEGALGSD 546
DB 8070 NEAQTRKVAEQQSQAALNDAMKHLRNSIQNQQSVRQESKYINASDAKQYINHAVREV 8129
QY 547 DNLDFAQNTVLDKDYVSKKILSSMKTRNKKEYIHYIQLQGDKISYEASCN--LFSKDPYS 604

DB 8130 ENINEQHPTLDKEII-KQLTDGVNQAND--LNGVELLDADQNAHQSIPTLMLHLNQAOQ 8187
QY 605 SILYQK-NIEGSETAYYYVADAIEKIDKYRPIQISNKRNIKULTFIGHGSEFNWTF 663
DB 8188 NALNEKINNNAVTRTEVAALIGQAKLLDHAMENLEESIKDKEQVK-----QSSNYINEDSD 8242
QY 664 ANLVDLSLSEIETILN-----LAKADISPKYIEIN-----LLGCNMFYSYIAEETYP 712
DB 8243 VOETVDNAVHVTIELNQTNPNTLSIEDIEHAINEVNOAKKQLRG-----KQKLYOTIDLA 8298
QY 713 GKLLLIKIDRVSELMPISQSDSITVSANGYEVINEEGKR-----EILD 756
DB 8299 DKLSKLDLTSQOSSSISNOIYTAKTREVAQAIEKASLNHAMKALNKVYNADKVL 8358
QY 757 HSGKWINKB-----ESIHKDISKEYISFNPKENKIIVKSKYLHELSTLLOEIRNN-- 807
DB 8359 --SSRFINEQPEKKAYQQAINHVDSIHRTQNPENDPTVINS--ITHELETAQNHLHGDQK 8416
QY 808 -----ANSSDIDLEKKVWLTECEIN-----VASNIDR-QIVEGREIEAKNLT 848
DB 8417 LAHAQOADAANVINGLIHLNVAQREYMINNTNATTREKVAKLNDAQALDKAMETLQQV 8476
QY 849 SDSINYIKNEFKLIESISDSLYDLKHQGLDSDSHFISFEDISKTEGPRIRPINKETGNS 908
DB 8477 AHK--NNILNDSKYLN--EDSKYQQQYDRIADAEOQLLQNTTNPTELEPKYKDIVKDNV--- 8530
QY 909 FIETEKIEIFSYATHISKEISNIKDTTFDNNVNGKLVKVNLDAAHEVNTLNSAFFIOSL 968
DB 8531 --LANEKILFG--AEKLSYDKSNANDEI-----KHMNYLNA----- 8563
QY 969 IEYNTTKESLSNLSVAMKVO--VVAOLFSTGLNTITDASKVVELVSTAL----- 1015
DB 8564 -QKQSIKMI SHAURTEVKKLQQAILEDAMKSLSDTKQV--ITDITLPNTYBASEDK 8621
QY 1016 DETIDLLPTLSEGLPIIATII--DGVSLG--AAIKELSETNDPL---LRQIEYEAKI--- 1064
DB 8622 KEKVD--QTVSHAQAIIDKINGSNVSLQVQALQSQALQSQALQSQALQSQALQSQALQSQAL 8678
QY 1065 ---GIMAVNLTAASTA--IVTSALGIASGFSILLVPLAGISAGISPLVNNELILODKATK 1119
DB 8679 TIDQLTHLSLQOQTAKESVKNATKLEB-----IATVSNNAAQNALNKWVGK 8723
QY 1120 VIDYFKHISLAEATEGAFTLLDDKIIMPODDLVLSEIDFNNSNITLGKCEIWRPAGSGSGHT 1179
DB 8724 LEQFINHADSVENSNYRQADDDKIAYDEALEHGQDIQKTNATQNETK----- 8772
QY 1180 LTDDIDHFFSPSITYRKFWLSIYDVLN--IKKEKIDFKDLMLVLPNPNRVFGYEMGWTP 1238
DB 8773 -----QALQOLYIAYETSLNGFERLNHARPRALEYIKSLEKINNAAQSALEDKVTQSH 8824
QY 1239 GFRSLD---NGTKLLDRIRDHYEQGFYWRVAFIADALITKLKPRYEDTNNRINLDGNT 1295
DB 8825 DLLELEHIVNEGTLNDIMGE-----LANAIVN---NYAPTKAISN----- 8862
QY 1296 RSFIVPVITTEQIRK-NLSYSFYGGSGYSLSLSPYNNNIDILNVENDTWIDVD----- 1349
DB 8863 -----YNAOLNRKDNFTQAINNARDALN---KTQGNQLDFNAI--DTFKDDIDFKTDA 8911
QY 1350 -NVKNITIEDEIQGELIENI--LSKLNI--EDNKIILNNHTINFYGDINESNRFISLT 1405
DB 8912 LNGIERLTAASKAEK--LIDSLKFIKAQTHANDEIMTNS-----IAQLSRIVNQA 8963
QY 1406 FSILEDINIIIEIDLVSYSKILLSGNCMKLIENSDDIOQKIDHIGFNGEHQKIFYYSYI 1465
DB 8964 FD-LNDAMKSLRDELNNAQFPVQASSN---YINSDLEDLQKQFDHALSNA--RKVL----- 9012
QY 1466 DNETKYNG-FIDYSKKEGLFTAESNESIIRNIYMPDSNNLFIYSSKD-LKDIRIINKGD 1523
DB 9013 ---AKENGNLDEKIQIGL-----KQVIEDTKDALNGIQRLSKAK 9049
QY 1524 VKLLAIGNYFKDMKVSLSFTTIEDTNTIKLNGVYLDENGVAQILKFMNNAKALNTSNSIM 1583

Db 9050 AKAI--QYVQ-----SLSYINDAQRHIAENNIH-NSDDLSSLANTLSKASDLDNAMKDRLR 9101

Qy 1584 NFLESINI ---KNIFYNLDPNIEFILD 1608

D**b** 9102 DTIESNSTVPNSVNYINADKNLQIEFD 9129

RESULT 9

AAB18172
ID AAB18172 standard; Protein; 2485 AA.

AC AAB18172;

DT 07-NOV-2000 (first entry)

plasmidium falciparum chromosome 2 related protein SEO ID NO:29.

plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW
antimalarial; malaria; protozoacide; infection; insecticide.
KW

Plasmodium falciparum.

PN WO200025728-A2.

PD 11-MAY-2000.

05-NOV-1999;

PR 05-NOV-1998; 98US-0107131.

PA (HOFF/) HOFFMAN S.

PA (GARD/) GARDNER M.

[illegible]

XX
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 104

XX
XX
12/15/2000 12:14 PM

PT
proteins encoded by chromosome 2 of the human
Plasmodium falciparum, useful as antimalarial vaccines and in the

[illegible]

CC The present invention describes proteins and their fragments (1) encoded

CC by chromosome 2 of the human malarial parasite, *Plasmodium falciparum*.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against *P. falciparum* infection comprising (I) or (II).

(I) and (II) are useful for the development of vaccines against *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are

CC useful in the detection of infection with *P. falciparum*. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent *P. falciparum*

CC energy, or they can be used to identify any relevance an
CC *P. falciparum*. Sequencing of the plasmidom chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand

CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito

CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide

CC specifically mentio

SQ Sequence 2485 AA;

Query Match 3.9%; Score 337; DB 21; Length 2485;

Matches 356; Conservative 267; Mismatches 646; Indels 540; Gaps 91;

Qy	35	HNMSSESVVEKYLKDINNLTDN-----YLNTRYKSGRNKALKXF-----KEYLTMEVLE	85
Db	157	HTNSNI---YINCNDNNNDIENSNVQHYNDKIKENINKQKQVYLINDINNKYI-	212
Qy	86	LKNSLTPVEKLNHFHWIGQINDTAIYNQ-----WKVNSDYTVKVPYDSNAFLN	139
Db	213	LSKNTCKINKGKL-----IKKKVNNISRNRHILYCRNK-----LYNGVNF---	257
Qy	140	TLKKTIVESATN--NTLEGFRENLN---DPEFDYKPYRKMBIYDKQKHFDYYSQI	194
Db	258	--SDDIIKSEVAVCNLSLTVLHKYNNIMDMNYLDN-----IHTNNSNIYD-----I	301
Qy	195	EENPEFIIDNIKTY-----LSNEYSKOLEALN--KYIEBSLNKITAN-----	235
Db	302	NYTNERVINSTCRYPIGNNTLSKDEVTKSSKINSLSFYDDIIN-VNKNDDIPILHDK	360
Qy	236	NGNDIRNLEKF---ADEDLVRLYNQELVER-----WNLAASDILRISMLKEDCGVY	284
Db	361	NINIIISKESCHDEKEBEKYIWNYNLVEEKQKQKIWN---SLNVLPIIDILKNG---	414
Qy	285	LDVDILPGIQDPLFKINPDSITNTSWEMIKLEAIMKYKEYIPGTVSKNFMDLDEEVQR	344
Db	415	-----HDEINKEIKCKKKKSPSQNDIKSKMLYNNKSY-----SKS-----	450
Qy	345	SFESALSSKDKSEIFLPLDDIKVSPLEVKIAPFANNSVINQALISLKDSCSLVI----	400
Db	451	--EKVLYTNNKNSNTPIPFPLN---KVGDKFKNSENIYDMYNNKQVYIHDKKIYTNN	504
Qy	401	--NQIKNRY-----KILNDNLNPSINECTDFNTTM--KIPSD-KLASISNDNNMMFI	448
Db	505	YSNKLQKHYYTSNINLLYNNIGKVDNLGHLHUSNNMYCRLNSNPPYKSI SLNNNNVFFY	564
Qy	449	KITNYLKVGFAPDVRSTINLSGPGVVTGAYQDILLMFKD-NSTNIHLLEPE---LRNPEFP	504
Db	565	K-----KXKSNNSNNNNNNISSSSSSSKKHVIINKKISSYNIHYKERKDSFKNFLFP	620
Qy	505	KTKISQLTQBEITSLMSFNQARAKASPEEYKGYFEGALGEDDNLDPQNTVLDKOVSK	564
Db	621	KEKI--LPKKOTCV-----FNERQKOLFE-----	643
Qy	565	KILSSMKTRNKEYIHYIVOLQ--GDKISYEASCNLFSKOPY-----SSILY---QKNIEG	614
Db	644	-----KSNEHIKCVSFPNTSDDISSHSVN--KKEPFPALONNSIRHIPPENNIY	693
Qy	615	SETAYYYYADABIKEDKYRIPQYSNKNKIKLTFIGHGKSBFTDTPANLDVDSLSE	674
Db	694	TSKGSFNHWQDEKTVLLKK--KEINDKNTFSSCLINHNIITYTTLONGVKN-----	744
Qy	675	IETILNAKADISPKYIEINLLGNMFSYI-----YAEETYPGKLLKIKDRVSELMPS	729
Db	745	-----LNMLGIRDSIYKIDKNNMLKECYNGNNDNSNNKKKKKKKLS	786
Qy	730	ISODSITVSANQYEVRIENEGKEIILDHS--GKWINKESIIDKISSKE-----YISFNP	782
Db	787	FSCDIIINDNITPYESDKENKNSNNIKSMDFINTVYKRSNLYNLSNRDSTVDMHKNYS	846
Qy	783	KENKIIVKSYLHELSTLQEIERNANSSDIDLEKKVMLTECBI--NVASINDROQIVEGR	840
Db	847	EYINTORTNKIYELSN--KRIRNYKLYS-MDEIFKVSLEKKEYIDNISNMER-----	897
Qy	841	IEEAKNLTSDSIYIKNEF--KLIESISDLYDLKHQGLDGHSHFISFE-DISKTENGFR	897
Db	898	-----VTY-KNEMINEKISKMDIILYPCDKNKS LNMSPCVIENNISREEN---	942
Qy	898	IRFINKETGNSIFIEYEK-----EISEVATHISKEISNIKOTIFDNVCK-----LVKK	947
Db	943	-----EKSSVTLNKKQENEMFNCVGRLCHMGK--MNNQDNIIYDQGNIKQNEEETKH	994
Qy	948	VNLDAAEVNTLNSA-----FFIOSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNT	1000
Db	995	DEYISREKKNYSKIRPNDDYKYQVLSYHTLDEDEKKK-----	1034
Qy	1001	ITDASKVVELVSTALDETIDLLPTLSEGLPIIATII-----DGVSLGAAIKELSETNDPLL	1056

Db 1035 -NDMNLIDMNEAIIETVN-----GVINNIIIDRKDNNSRDKMEKEMEKEMKQM 1084
 QY 1057 RQIEAKIGIMAVNLTAATAIVTSALGTASFSILLVPLAGISAGIPSLVNNELIQDK 1116
 Db 1085 EKEMEK---VMEKEMEKMEKEVEKEL-----KNEMNNMNNMNNMNNMNNMNN 1129
 QY 1117 ATK-----VIDYFKHISLAETGAFTLLDDKIIMPDDVLSEIDFN--NNSITLGRKEIWR 1171
 Db 1130 IYKNEIYVNDNDKEIIVNEE-----KKLIYFPN--YESDVHRKNNMSIINNCK--- 1177
 QY 1172 AEGSGHTLDDIDHFFSSPSITYRKPLWSIYDVLNIIKKEIDFSDMLVPLNAPNRVFG 1231
 Db 1178 -----DDYN-----NILKEYVDS-----CLAQKEENI-- 1200
 QY 1232 YEMGWTPGRSLDNDGTCLLDRIDHYEQFYWRYPAP-----IADALIT 1276
 Db 1201 -----FRPLFNKK--DKV-----WKRENIKNNTIITHNEEMKRIYQTINK 1241
 QY 1277 KLKP-----RYED---TNVRINLDGNTRSFIVPITTEOIRKNLSYFSGSGSYLSLS 1328
 Db 1242 NVPPIYFNRYENFLNHLTYNPPKNDLPKLSYKVMNNIR-NLYTANGHNNNDYMK 1300
 QY 1329 PYNMNI-----DLNLVENDTWI-----DVDNVVKNITIE----- 1358
 Db 1301 LYNQNIYTLKYQVANIDNDHICKGGGLDYINMNIISKECKNRKDKTYLNKIFHYKKKD 1360
 QY 1359 -----SDEQKGBLNIENILSKLNIENDKILNNH--TINFYGD-----INESNRFISL 1404
 Db 1361 ARPFINDEIGSNIDYDIKKYSNDENNYKLNKNNISMSNDRDMLPTLNSEHNNFPSC 1420
 QY 1405 TFSILEDINIIBIDIVSKYKILLGNCMKLIENSSDIQOKIDHIGFNGEHQKIFYSY 1464
 Db 1421 QPNLEKKSYIDNLNYSNMDDFTEBKYNFVNNENDL-----FNTKRWKFNFSK- 1471
 QY 1465 IDNETYNGPIDYSKEGFTABFNSIESIRNTPDSSNNLFTYSSKOLKDRIRINKGVD 1524
 Db 1472 -GKNLFNNKFFVNSNEDGVFSF-FKMNLPREL-NKSNLSKLESYKSNNSNNCSNKGDD 1528
 QY 1525 KLLIGNY-----FKDMKVSL-----SFTIEDNTI--KUNG---VYLDENGA 1563
 Db 1529 N--IGNMENNMTTNVTIASDEHISTKGDIDHESFRDNDICILLKTEGRSKYSD----- 1581
 QY 1564 QILKFNNAKSALNTSNLNFLESINIKNIFVNNLDPNLEFIDTNTFIISGS-NSIGOF 1622
 Db 1582 --ITLNEKSNLNDNETINEYENV-----CSNIDVN-----EWEDKVGTCNSVGDK 1628
 QY 1623 ELICDKDKN 1631
 Db 1629 ETEKNNEKN 1637

RESULT 10
 AAB18222
 ID AAB18222 standard; Protein; 2496 AA.

XX AAB18222;
 AC AAB18222;
 XX 07-NOV-2000 (first entry)
 DT XX
 DE XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:79.
 DE XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW XX antimalarial; malaria; protozoacide; infection; insecticide.
 XX
 OS XX Plasmodium falciparum.
 PN XX W0200025728-A2.
 XX
 XX 11-MAY-2000.
 PD XX
 XX 05-NOV-1999; 99WO-US26796.
 PF XX
 XX

PR 05-NOV-1998; 98US-0107131.
 XX (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX WPI; 2000-365347/31.
 DR
 XX
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection -
 XX
 PS Disclosure; Page 185-192; 577pp; English.
 CC
 CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (I) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasitic lifecycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in new
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.
 XX
 XX Sequence 2496 AA;

Query Match 3.6%; Score 314.5; DB 21; Length 2496;
 Best Local Similarity 18.8%; Pred. No. 1e-07;
 Matches 370; Conservative 316; Mismatches 730; Indels 549; Gaps 96;
 QY 57 DNYLTYYKSGRNKALKPKF---YLTMEVLELKN--NSLTPVEKNLHPITWGGQINDTA 111
 Db 24 DNKSVLRSTTKKKKKICKSTFYVHBESEIKSWLRNSNRDRKGFPIFERLIKERR 83
 QY 112 INVINQWKVNSDYTVKFVYDSNAFLINTLTKTIVESATNTNLTSPRENLDPEFDYKPF 171
 Db 84 YICVKNYKRNK---LKWIYK-----NTYEKT-----KNICO---DYNLL 117
 QY 172 YRKRMEIIVDKOKH-FIDYKYSQIBENPEFI---IDNIIKTYLSNEYSKDLBALNKYIBE 227
 Db 118 PKCIGIIVDKNKEFTFFETFFENFDNIFYNNKYIFNIY--YMFEDYTKKKVKKIEGIE 175
 QY 228 SLNKITANGNDIRNLEK---FAEDDLVRLYNQELVERWNLAASDILRISMLKEDGGYV 284
 Db 176 NNNRHNNNNNNIIFYVHKFFLFNDDEKKKRDNDI--KINIKLHNNTRKLSVSEE----- 228
 QY 285 LQVDILPGTOP-----DLF-----KSNKPDSTINTSWEMIKLEAIKYEYIPG 329
 Db 229 -NVEUKPIYKQGERNETVNNLYEYTGGVKRNNNNEIIVVTSTEFHRIIVICFRTVGH 287
 QY 330 ---YTSKNFMDLDEEVQVSFESALSSKDSKSIIFLDDIKVSPLEVKIAFANNVINO 386
 Db 288 SHIITSPH-DALNHIVE-----ENDKIKLSEIYSIPFYIYGNLGLKN--VITTG 335
 QY 387 LISLKDSCYSDLVIN-----QIKNRYKIINDN---LNPSINEGTDFN 425
 Db 336 IVEFMPYPSRTQMNFTVTGANGEMNDLYKFDLKIIRIRIPRNTKKILGLSTNE----- 390

CC vaccines against *P. falciparum* infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against
 CC *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with *P. falciparum*. Furthermore,
 CC (I) especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent *P. falciparum*
 CC infection, or they can be used to identify drug resistance in
 CC *P. falciparum*. Sequencing of the plasmidium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasitic lifecycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.
 XX
 SQ Sequence 3973 AA;

Query Match 3.6%; Score 313; DB 21; Length 3973;
 Best Local Similarity 18.5%; Pred. No. 2.2e-07;
 Matches 375; Conservative 335; Mismatches 708; Indels 611; Gaps 90;

QY 11 KMVYV--KPIQED-----EVAIL-----NALE-----EYNNSESSVVEKYLK 49
 DB 808 KIYILRKQNKNDITYIYIEYILLDYFKYFYNNMDNYFKYNNMKSTVRKYVKN 867
 QY 50 KDINLTDNLYTKSGRNGKALKFKPEYLTWVLELKNLSLTPVEKNLHFIWIGQIND 109
 DB 868 KNIND-----TKGNK-----NNNN-----NNNIY-----GDGDD 892
 QY 110 TAINYNQKDV-----NSDYTKFYVDSNAFL-----INTLKTIVESATN 151
 DB 893 NNI-YCHDDDDIYCHDDDDIYCHDDDDIYCHDDDDIYCHDDDDIYCHDDDDIYCHDDDD 951
 QY 152 NTLESRENLDPEPDYKPYRKRMEIYDKQKHFD-----YKSOJEEENPEFTI--- 202
 DB 952 -----YFK-----IPYDKNTFINLTYLFFPSLKVQBFSSRQIMIV 989
 QY 203 -----DNLIK-----TYLSN-----EYS----- 215
 DB 990 RCLSFLLKKKINKLNAYIFQLFSYLENDQININEKGMVHRKSSKYHRRNQEEYSHNKT 1049
 QY 216 -----KDEALNKYIEESLNK-----ITANGNDIRNLEKFADEDLVLYNQELVE 261
 DB 1050 NDNSVSNLYRDIE--NEYDENHLERRKDRNVFSSNNMDDKYNLSDFKYTK-ENMDIKE 1106
 QY 262 RWNLAASDIIRISMLKEPDGVYLDVILPGIOPDLFKSINKPDSITNTSWEMIKLEATM 321
 DB 1107 NFRIDIS--FLKIFFLNDVR-----QINLNSGRKDKLESKAKRRIQKLDVH 1153
 QY 322 KY-----KEYIPGYSKNFOMLDEVOVSFESALSSKDSKSEIFPLPDDIKVSPLEVK 374
 DB 1154 RYTYENDNNKNDGNTF--LSQDEKSKSPDSDSCSDEKE----- 1195
 QY 375 IAFANNSVINQALISKDSYCSDLVINQIKRYKILNDNLNLSINSGTDFNTMTKIFSDK 434
 DB 1196 -----SSKGLYGNDFVNSDDHNNSSNNSSNNSSNNSSNNSSNNSSNNSSNNSS 1241
 QY 435 LASISNEDNMFMK-----ITNLYKVGFPDVRSTIN-----LSGPGVYTCAYOD 480
 DB 1242 VVDPDYDNYECKDSNKGVVVNYF-YAHLFPNFEKSYNNYVVEDISFDDIFLISIMD 1300
 QY 481 LLMFKDNSTNIHLLEPELRFNFPKTK-----ISQTEQEIITSLMSFNQARAK 528
 DB 1301 LWETNNNNLLNLNDLLKTYEEKKKIYICTSLLLKIFKRIIKKSSSYFLFYIYA- 1359
 QY 529 SOFEYKKGYFEGALGEDNDLPAQNTVLDDKYVSKILSSMKTRNKYIHHIVLQGDK 588
 DB 1360 -----FENDI-----KLIDSINILIKKVVW----- 1381

QY 589 ISYBASCNLFSDKPYSSILYQKNIEGSETAYVYVADAEIKEDIKYRIPYQISNKRNIK 648
 DB 1382 -TFKCDNIFNRE--KNINIKLVKLFISYKLVKNYFLQIYYHFFYNNQIYNERKNYF 1438
 QY 649 -TFIGHGSEFNTDTFANL--DVSLSSEIETILNLAKADISP-----KYI 691
 DB 1439 DNFFSIFSKYINKIFVBIYSSSSSTSSSFVFNVSFKYMMKMCISIIINNMIGWKYI 1498
 QY 692 EIN-----LLGCMSESYSIYAETYPG-----KLLLKIKORVSELMPSI 730
 DB 1499 NLERVKQVYEHINIMDMVHMSHLHDIDVYGHDSNNYNIYQKIISKYGEKDTLDVI 1558
 QY 731 SODSITVSANQVEVR-----INEGKRBIIDHSGKWINKESIIKDISKEYISFNP 782
 DB 1559 NTESVHQNRNEDDIDGSLNSLDVFNEMRNIIINH-----SLIKD----- 1599
 QY 783 KENKIIVSKYLH--ELST--LLQIRNANS-----SIDLEKKVMLTECEINVASNDR 834
 DB 1600 -HNDMCTKKGRINIFQISSPATSEQLMNNHYTMNYLTDVMLLQKDYIYNDNM--NEHK 1656
 QY 835 QIVEGRIBEAKNLTSDSINYIKNEFKLIESDSLY--DLKHQGLD--DSHFISFEDIS 890
 DB 1657 QNVFNKPPDNN 1716
 QY 891 KTENGFRIRFINKETGNSIFITETEKEIFSEYATHISKEISNIKOTIFDNVNGKLVKYNL 950
 DB 1717 NEENELIQ--NNSMSSSIYI--DKLMDK-----TKEMEPLFNKTKDMKNYNEOKNE 1767
 QY 951 DAAHEVNLT--NSAFFIQ-----SLIEYNTTKESLSN--LSVAMKVQVYAQLFSTG- 997
 DB 1768 LISYPYNNMLQNNIIFVKFFLYTQNLQIIFQNNYIFFLSDFLFNYKKKEYIEBKNGN 1827
 QY 998 --LNTITDASKVVELVSTALDETDLPLSEGLPIATIIDGVSGLGAIKELSETNPL 1055
 DB 1828 QNVINIKEDKHI-----TNIKDGDKHTNIKDGDKNTNIKDDDKNTNM 1873
 QY 1056 LRQETEAIGIGAMVNLTAATAITVSALGIASGFSILLVPLAGISAGISLVNNEILLQD 1115
 DB 1874 KKNNKNYLTILMYSQESFY-----YSI-----FNTLINDYFLAY 1911
 QY 1116 KATKVIDYFKHISLAETEGATFLD-----DKIMPQDDLVUSE--- 1154
 DB 1912 K-----DYKSCFLYESLNTFFKINNFNNYFLCKYSSGYLPLERIIKLFMDVIFCHFTK 1966
 QY 1155 -IDFNNSITLCKEIRAEAGSGHGLTDDIDHRESSPSITRYKPWLSIYDVLNKKKI 1213
 DB 1967 FININENINDVELLEVLYNGKCYELLRFL--FF-----FIKONDTITINIKY 2014
 QY 1214 DFKDLMVLPNAPNRFVGY--EMGWTGFRSLDNDGTKLLDRIDHYEGQFYRYFAP 1270
 DB 2015 IFDIIMCILEQYIAHVNYIYKKAWDVFFNKL-----KILN--LSLHFNYSIYFIK 2068
 QY 1271 ADALITK-----LKPREDYTVNRIINDGNTSRFIVPVITEQIRKNLSYS---- 1315
 DB 2069 INAEIKRENDNNKSNNDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2128
 QY 1316 -FY-----GSGGSY-----SLSLSPYNNVINDNLVENDTWIDVDN 1350
 DB 2129 LFYNNVINKLHVNSINCLKNKLGYSKDEEPPKELNCKRYFLNY-----NKDFKKEIY 2182
 QY 1351 VYKNTIETESDEIQKELIENILSKLNIEDNKIILNNHTINFYGDINENRISLTFSTILE 1410
 DB 2183 VLYNLIAS-----EIFELIKAIYINETKIY--PLINICYDRNISINFINIDY--D 2231
 QY 1411 DINIITIE-IDLVSKYSKILLSGNCKMLJENSDDIOOKI-----DHIGNGEH---QKYI 1460
 DB 2232 NLNSILEKYTVLHKKKDKHINKLVLLCKKSIHMHKVIYSIDDDHLLNNMLHLRRKNI 2291
 QY 1461 FYSYIDNETKNGFTDYKKEGLFTAEPNSESIIRNIYMPDNNLFIYS-----KDL 1513
 DB 2292 YKYYVLNINEYNNFLDNHKKCR--KRKFYNNYNOSSY---NNNNYINNTNTNFFEYHDY 2346
 QY 1514 KDIRIINKGDVKLLIGNYFKDDMKVLSFTIEDTNTIKNGV--YL-DENGVAQILKPMN 1570

Db 1436 NNNKFMICNHEQDTE---EKGNTQNKERKHDIVIGPIYNNSPDSTTTTHSSNNYKGNH 1492
Qy 789 VKSKY-----LHELSTLLOE-----IRNANSSDIDLEKKVMLTECEINVASNIDROI 836
Db 1493 VSGDYKNDGLLHGNNMNECYVKOTKCNVNNNNNNNNNNNNNNNNNNNNNNNNNNNNND 1552
Qy 837 VEGRIEAKNLTSDSI---NYIKNEPK--LIESISDLYDLKHQGLDD---SHFISPD 888
Db 1553 I---IKYHTIDTNSKNHTYFNKFLNFDKLIISNIYGLPQGFSLSNILCSLYAYLD 1609
Qy 889 ISKTENGPRIRFINKTGNSIFIETEKEIFSEYATHISKEISNIKIDTIFDNVNGK-----943
Db 1610 --KNEEFQNLVSEKQINNKYFLANGT---CNVFNLSLILRFIDDFLITLKNKNIKIF 1664
Qy 944 ----LVKKNLDRAHEVNLNSAFFIOSLI-----EYNTKESLSNLSVAMKV 987
Db 1665 KNLLKKKI---WGSNINSKTKIPKIPLIYKNDLLIYNFQNKYQKKYKIKNKKKIQSV 1721
Qy 988 ---QVYAQLFSTGLNITDASKVVELVSTALDETIDL--PTLSEGLPIIA--TIIDGVSL 1041
Db 1722 RKRHIHQLVANKKHSTVQK-----DKINKYINLIHPTIQKNDVLSNNSIMNPERI 1775
Qy 1042 GAAIKELSETNDPLLRQEIEAKIGIMAVNLTAATAVTSALGIASGFSILLVPLAGISA 1101
Db 1776 --YIKESHKSN-----SSIRT 1789
Qy 1102 GIP-SLVNNELLQDVKATKVIDYFKHISLAETEGAFITLDDKIIMPQDVLVSEDFNN 1160
Db 1790 DIPNSVNDIDIEYNQSD-----NNSYSTNNLYNNINMTQNG-----DNN 1829
Qy 1161 SITLGKCEIWRABGGSGHTLTDDIDHPFSSPSITYRKPLWSLYDVNLNKEIKIDFSKDL 1220
Db 1830 NVNIFK-----HVQNSQFCFNSN-----NLYEKDIRENNI-----1861
Qy 1221 VLPNAPNRVFGYEMGTWPGFRSLDNDGTLLDRIDHYEGQFYWRYPAFIADALITKLP 1280
Db 1862 ---SQINRKLCKSRNFTKSRKINTLTYLQDKV-----IKILCKK 1900
Qy 1281 RYEDTNVRINLGNTRSFPVUPTTEQIRKNLSYFYGGSGYSLSLSPYNNI--DLNVL 1339
Db 1901 KYIKHIKKMKYMNQNF-----KKLKKQFP-----HNASFELKINKINKNIRLNKL 1949
Qy 1340 ENDTWIDVDNVVKNITIESDEIQKGLIENILSKLINIEDNKIILNHTINFGVDINES 1399
Db 1950 KKR-----KNHSI-----NTPVTSIE---WLNNSYTFDF---INNS- 1980
Qy 1400 RFISLTFSLIEDINIIIEIDLVSYSKYILLSGNCKMLIENSDDIOOKIDHIGFNGEHOKY 1459
Db 1981 -----IQSTSYP--WKNKCDATIRNHLHNV-----IDKNKNTY 2014
Qy 1460 IFVSYDNTKNGFYDYSKKEGLFTAEPFSNESIIRNIYMPDSNLIPIYSSKDLDIRII 1519
Db 2015 FMKNLVENRIVRNIISKQKQCOSLYKNK-----QNVYFCYKNPFLSKSLKFCICI 2067
Qy 1520 NKGDVLLIGNYFKDMDKYSLSFTIEDTNTIKLNGVYLDENGVAQILKPMNAKSALNTS 1579
Db 2068 -----KTLKMPNAFT-----NSTY-----NTKFL-----2088
Qy 1580 NSLMNPLESINIKNIPYNNLDNPIEFILDTNFTIISGNSIGOPELICDKDKNIQPYFINF 1639
Db 2089 -FLISYMNKMLIKN-----KKLKEV--KLFLIQTAEAPRYARIFNOODSFPYCLQHF 2138
Qy 1640 -KIKETSYLY-VGNRQNLIVERSYHLDSDGNSISSTVINFPSQKLYGIDRYNKKV 1692
Db 2139 RKIKKRLINKYKIGHKNLLREFF-----LNFNIKELYNWPYMFKI 2182

RESULT 13
AAW54145
ID AAW54145 standard; Protein; 1639 AA.
XX
AC AAW54145;

XX 25-MAR-2003 (updated)
DT 23-SEP-1998 (first entry)
XX P. falciparum synthetic gpl90 protein.
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite.
OS Plasmodium falciparum.
Synthetic.
PN WO9814583-A2.
XX 09-APR-1998.
XX 02-OCT-1997; 97WO-EP05441.
XX 02-OCT-1996; 96DE-1040817.
XX (BUJA/) BUJARD H.
PI Bujard H, Pan W, Tolle R;
DR WPI; 1998-240088/21.
DR N-PSDB; AAV21451, AAV35363.
XX Recombinant production of complete gpl90/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content
XX Example 1; Fig 3c; 48pp; German.
XX This sequence represents a modified Plasmodium falciparum gpl90/MSP-1
CC (merozoite surface) protein. The gene encoding this protein has been
CC stabilised by reducing the AT content of the nucleotide sequence. Such a
CC protein is useful in vaccines against malaria or for producing monoclonal
CC antibodies (for passive immunisation). The complete gpl90 protein can now
CC be produced outside the parasite and has, at least over extended regions,
CC the native pattern of folding. Larger amounts of the protein can be
CC produced recombinantly than would be possible using the parasites as
CC source.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX Sequence 1639 AA;
SQ
Query Match 3.6%; Score 309; DB 19; Length 1639;
Best Local Similarity 19.5%; Pred. No. 1.1e-07;
Matches 351; Conservative 280; Mismatches 554; Indels 618; Gaps 94;
Qy 36 NMSESVVEKYLKLDINNLTNDYNTYKSGRNKALKKFKLEYLTWEVLEKQNSLTPVE 95
Db 112 NPSDSSSDSAKSYADLKHRVRYLLT-----IKELYPQLFDLTNHLMLTCLD 159
Qy 96 KNLHFIWIGQINDTAINY-INQKDVNSDYTVKPYVDSNAFLNTLTKTIVESATNNTL 154
Db 160 -NIH-----GFKYLDGYEIN-----ELYKLN-FYDLLRAKLVNCANDVC 201
Qy 155 E-SFRENLDPEFDYNNK-----FYKRMBEIIYDKQKHFDIYKQSIENPEFIIDNIKIT 209
Db 202 QIPPNLKRANELDLVLKLVFGYRKPDLNIDKNVGMEDYIKKN-----245
Qy 210 LSNEYSKOLEALNKYIEESLNKITANNNDIRNLEK-----FADELIVLYNQELVERNVL 265
Db 246 -----KKTIENTNELIEES-KKTIDKNKNATKEEBKKLYQAOYDL-SIYNQOLBEAHL 298
Qy 266 AASDILISMLKEDGGVYLDVDILPGIQ-PDLFKSINKPDSITNTSWEMIKLEAIMKYK 324
Db 299 ISVLE-KRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNL-----LDKNKKIE 348
Qy 325 EYIPGYTSKNFMDLDEEVQVSFESALSSKSDKSEIFLPDLDIKVSPLEVK--IAFANNSV 382
Db 349 EH-----EKEI-----KETAKTIKFNIDSFTDPLEEYVLRKKNKI 386

Qy	383	INQALISL	KDSY	CDLV	INQIN	--RYKIL	NDNL	NPSI	NETD	FN	TTWK	IFSD	KLAS	ISN	440																								
Db	387	DISAKV	ETKES	----	TEPNE	YPNG	VTYPL	STYND	INNAL	NELNS	FG	----	----	----	427																								
Qy	441	EDNMFM	IKITNYL	KVGP	APDVR	STR	STIN	LSGG	VVTG	AVQD	LLMP	KD	NS	TNHL	LEPEL	RN	500																						
Db	428	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	451																						
Qy	501	FEFPKT	ISQI	TEQ	--EIT	LSW	FNQAR	AKSOF	EYKKG	YFEG	ALGED	DN	LD	FAQNT	VLDK	559																							
Db	452	F-----	INEIK	EIKI	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	486																						
Qy	560	DYVSK	KLSSM	----	KTRN	EYI	HVIV	LOQ	DKIS	EYAS	CNLF	SKDP	YSS	ILYQ	K-NTEGS	615																							
Db	487	EY--EKL	NEI	YDSK	FN	NDIL	TNFK	QMG	KRY	SYKV	E-KL	THNT	FAS	YSKHN	LE--	541																							
Qy	616	ETAYYYY	VADAE	EIKEID	KYR	PIYOI	SNKR	NI	KLTF	IGHG	KSE	FN	TDF	FANL	DVDS	LSSEI	675																						
Db	542	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	578																						
Qy	676	ETILN	LAKAD	ISPKY	EINL	GCNM	FYSI	YAE	TYPG	KL	LK	IKDR	VSEL	MPSI	SQDSI	735																							
Db	579	ETLVEN	IKKDE	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	615																						
Qy	736	TVSANQ	YEV	RINE	BKREI	DHSG	KWINK	EESI	IKDIS	SKEYI	----	-----	-----	-----	-----	-----	787																						
Db	616	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	664																						
Qy	788	IVKSKY	LHEL	STL	QET	RNN	ANS	SDI	DLEK	KVML	TECE	IN	VAS	NI	DQI	VEGRI	841																						
Db	665	IVLKEI	DKLV	FPK	VE	S	LINE	-----	-----	-----	-----	-----	-----	-----	-----	-----	715																						
Qy	842	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	888																						
Db	716	KPGQ	AGS	ALR	GSV	QAQ	QKQ	APVP	PVPV	PEAK	QAQ	PTP	PAPV	-----	-----	-----	763																						
Qy	889	ISKTE	NGFR	IR	FINK	--ETG	NSI	EIE	TEKEI	FSE	VATHI	SK	EIS	NI	KDTI	FDN	VNG	KL	946																				
Db	764	NKTEN	VS	KDYL	EKL	EFL	NTSY	--CH	YIL	VSHST	-----	-----	-----	-----	-----	-----	-----	-----	805																				
Qy	947	KVNLD	AAH	EVNT	LN	SAP	FIQ	SILQ	IBY	NTTK	ES	LS	NLS	VAMK	VQY	VAQL	FST	CL	NTIT	DASK	1006																		
Db	806	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	820																		
Qy	1007	VVELV	STAL	DET	ITD	LLPT	LSGL	PI	ATID	GV	--S	IGAA	I	KE	USE	-----	-----	-----	-----	-----	1053																		
Db	821	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	870																		
Qy	1054	PLLRQ	EAK	IGI	MA	VNL	TA	ASTA	IV	SAL	GI	SGP	S	ILL	VPL	AGI	SAGI	PSL	--V	NNELI	1112																		
Db	871	KIKNL	LEEAK	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	914																		
Qy	1113	LQDKAT	KV	IDV	FKHI	--SLA	E	TEGAP	--TIL	DDK	I	IMP	OD	DL	VLS	BEID	--F	NNNS	I	T	LGK	CEI	1169																
Db	915	HSTNL	NSL	KL	FEN	IL	S	LGK	KN	YI	QELI	G	K	S	SEN	F	EK	L	K	D	S	D	Y	NP	970														
Qy	1170	WRASG	GGSH	T	DDI	DH	FFS	P	S	T	Y	K	P	W	L	S	I	--Y	D	V	N	L	T	K	E	ID	--F	S	-----	-----	-----	1216							
Db	971	VKSAD	DINS	LN	DES	K	R	K	L	E	D	I	N	K	L	T	L	Q	L	S	F	D	L	Y	N	K	L	R	L	F	D	K	K	T	Y	G	V	K	Y

QY	1339	-----VENDTWVI--DVNDNVVNKNTI-----ESD	1367
Db	1191	EGTDVATVWVSSESGSDTLEQSQPKKPASTHVGAESNTITTSQNDVDEYDDVIIVIPFGES	1250
QY	1361	E-----IQGE-----LLENILSKLNIE-----DNKIILNNHTINFYGD	1394
Db	1251	EDYDDLGQVVTGEAVTSPVIDNILSKTENEYEVLYKPLAGVYRSULKQLENNVMTFNVN	1310
QY	1395	INESNRFSISLTFPSILEDINIIIEIDLVSYSKILLSGNCM-----KLIENS	1440
Db	1311	VKD---ILNSRFKNRFKNVLESDLI--PYKDLTSSNVVVKDPYKFLNKEKDKFLSSY	1365
QY	1441	SDIOQKID-HIGFNGEHQKIFYSIDNETKNGFDYDSKEGGLTFAEPNESIIRNIYM	1499
Db	1366	NYIKSDSDTDTINFANDVLGY--KYKLSL--KYKSDLD-SIKKYINDKQGENEK-----YL	1415
QY	1500	PDSNNL-FIYSS-KDLADIRIINKGDVKKLIGNFYKDDMKVLSL-----FTIEDTWTIKL	1552
Db	1416	PFLNNITLTKTVNDKIDLFVIHU-EAKVLNYYEKSNEVEVKELNLYLKTIOQ----	1470
QY	1553	NGVYLDEN--GVAQILKPMNNAKSALNTSLSLNFLESINI-----KNIPYNNLDPNIEF	1605
Db	1471	ADPKKNNNFVGIADL-----STDYNNHNLTKFLSTGCVFENLAKTVLSNLLDGNLQ	1523
QY	1606	ILD 1608	
Db	1524	MLN 1526	
RESULT 14			
ID	AAE29345		
XX	AAE29345	standard; Protein; 1639 AA.	
AC	AAE29345;		
XX	27-JAN-2003	(first entry)	
DT	Plasmodium falciparum merozoite surface protein-1 (MSP-1).		
DE	Band 3 polypeptide; malarial infection; drug resistance; vaccine;		
KW	protozoacide; gene therapy.		
XX	Plasmodium falciparum.		
OS	WO200270542-A2.		
PN	01-MAR-2002; 2002WO-US06415.		
XX	12-SEP-2002.		
PD	02-MAR-2001; 2001US-272930P.		
PF	(SELI-) ST ELIZABETH'S MEDICAL CENT INC.		
XX	Chishti AH, Oh SS, Liu D, Goel V;		
PI	WPI; 2002-759814/82.		
DR	N-PSDB; AAD46980.		
XX	New isolated Band 3 polypeptide which selectively binds to merozoite		
PT	surface protein-1, useful for the prevention and treatment of malarial		
PT	infection		
XX	Disclosure; Page 113-118; 163pp; English.		
PS	The invention relates to an isolated Band 3 polypeptide that comprises		
XX	any of 4 20 residue amino acid sequences, or their fragments that bind		
CC	to an MSP-1 polypeptide or a polypeptide with any of 8 291-1331 base		
CC	pair sequences. The methods and compositions of the present invention		
CC	are useful for the prevention and treatment of malarial infection. The		
CC	present invention, develops new and more improved methods based upon		
CC	inhibiting the particular interactions between the malarial parasite and		
CC	a cognate molecule present in the host and subsequently minimising		

Disclosure: Page 113-118; 163pp; English

The invention relates to an isolated Band 3 polypeptide that comprises any of 4 20 residue amino acid sequences, or their fragments that bind to an MSP-1 polypeptide or a polypeptide with any of 8 291-1331 base pair sequences. The methods and compositions of the present invention are useful for the prevention and treatment of malarial infection. The present invention, develops new and more improved methods based upon inhibiting the particular interactions between the malarial parasite and a cognate molecule present in the host and subsequently minimizing

KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoicide; infection; insecticide.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200025728-A2.
 XX
 PD 11-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26796.
 XX
 PR 05-NOV-1998; 98US-0107131.
 XX
 XX (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX WPI; 2000-365347/31.
 DR
 XX
 XX Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection -
 XX
 PS Disclosure; Page 120-124; 577pp; English.
 XX
 CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)
 CC vaccines against P. falciparum infection comprising (i) or (ii).
 CC (i) and (ii) are useful for the development of vaccines against
 CC P. falciparum infection. (i) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (i), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (i) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasitic lifecycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.
 XX
 SQ Sequence 1516 AA;
 Query Match 3.5%; Score 306.5; DB 21; Length 1516;
 Best Local Similarity 18.2%; Pred. No. 1.4e-07;
 Matches 311; Conservative 278; Mismatches 494; Indels 627; Gaps 81;
 QY 6 KAOLQKVVYVFRIOEDEVATLNALEEVHNMSSSVVEKYLKLDINNLTQ-----NYL 60
 DB 131 KTAELKVTNYQ-----RTLNSMKSNNKNDNS-----NNIEDKTPNKT 173
 QY 61 NTRYKSG-----RNKALKKFEYL-----TWLE 85
 DB 174 NTQNSKNTQNSKNTPKINADISKSLIQIYDDIKEKOKSLNSLVHVGNVFVPSVKDVL 233
 QY 86 LKQNSLTPVERKHLFI-----WTGGQINDT-AINYINQKDVNSDYTVKFPVDSN-AP 136
 DB 234 ICNDDLSKIKKIFMTDGPVFLFGQDGMGTVENIKLDRNKD-----ENNLISY 286
 QY 137 LINTLKKTIVESATNTLESRENLENDPEFDYKFKRMEIYDKQHFIDYKSOIEE 196
 DB 287 SINYKQVDVNNNDKDKKKNENVRD-----QKNYV--YKKN--- 326
 QY 197 NPEFIIDNIITLSNEYSKDLALNKYIEESLNKITANNNGNDIRNLEKFADEDLVLRL-- 254

DB 327 -----ENINNYLDDDEK-----EDIQNGVYNNDDI-----DEQIRKGH 364
 QY 255 -----YNQELVERWNLAAASDILRI--SMLKEDG 281
 DB 365 MARKKYIESIPKTPKGFCLMRPVDIIDIISNTNTEMLR-----ISETLKUHENFKQHL 418
 QY 282 GYLDVDILPGIQDPLFKSINKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKFNMDLDBE 341
 DB 419 NVLDENNSTPVVNNLLKNIN-----YK-----KNDDLIEGG 450
 QY 342 VQSPF-----ESALSSKSDKSE-----IFLPLDDIKVSPLEVKIAPANNSVIN 384
 DB 451 EKKSFINLINVDSCYSSNSRLNENDENIERGKINNFITNDEKSIINN-----YNNNNNN 506
 QY 385 QALISLSDSYCSLDLVINOIKNRYKILNPLNPSINEGTFNTMTKIFSDKL--ASISNED 442
 DB 507 NNNNDNNNDNDVIIENKNNNNIYDNKN-----VECSSEKINDNGISNKN 554
 QY 443 NMFMIKITNYLKVGFPADVRSTINLSGPGVYTGAYQD-----LLMFKDNSTNIHLLEPE 497
 DB 555 -----INILELPNNLDT-----SNIFLEKDEYKVYVYVANKKEIRIPLFKKEINKEI----- 599
 QY 498 LRNFPFPKTKISQLTQEQITSIW-----SFNQARAKSQFEBYKGY--- 538
 DB 600 ---FEKLPKLQYQILQDIKEBWTNDRIKAISKDDMDVFSQVQLETYVRMIKTDFFE 656
 QY 539 -----FEGALGEDDNLQPAQNT-----VLDKDYVSKKILSSMKTNRKEIHYIV 582
 DB 657 KLIKMAENIOSVEGEL--LINKOLSKNTNINIKDYNVLQKK---KSKKKKKFLNDIL 711
 QY 583 QLQGDKISYEASCNLFSDPYSSILYQKNIEGSETAYVYVADABIKEDIKVRIPOISN 642
 DB 712 N-----TYNFTTSKYQDL-----YVKGESKEDIKNQIDFVTQE 746
 QY 643 -KRNTKLTFIGHKGSEFNTDFANLDVLSLSEIETILNLAQADISPKYIEINLGLCNMF 701
 DB 747 CYRNNDIIRDTHDKS---DIFQNIKIDNNKKYIYNLELEOEINEKK-----NYN 794
 QY 702 SYSIYAEETYFGKLLKIKDRV-SELMPISIQ---DSITVSANOYEVRI---NEEGKRE 753
 DB 795 KNNDSNKTFF---FLKIENFKDLDLDDSQIFGDSLLADIKEYNYTADNLDNNENKS 850
 QY 754 ILDSHGKWKINEESIIDISKE---YISFNPKNK--IIVKSKYLHELSTLQIBIRNA 808
 DB 851 LYEDGENFITRNEPITNEVEEKNNIYISDEQYNEEDIIFKDK-----IKERKQND 903
 QY 809 NSSDIDLEKKVMTCEINVASNIDROIQVIEEAKNLTSD-----SINYIKNEFKLIES 864
 DB 904 TSSD-DFE-----NCSVQ-----EKIYVNEKIEEYNNKNDKSSSSSSIILEEIKYKE 951
 QY 865 LSDSLY-----DLKHQGLDSDHP----- 883
 DB 952 KKDELVSPNLCVLLDDEFESNDLENNYISVSSDDMTKNTVKNKNTGVKENVKDKTNVEYD 1011
 QY 884 -----ISFEDISKTEGFRIRFINKECTNSIF---IETEKEIFSEYATHISK--- 927
 DB 1012 KKGDDGVIEISFDSHKLLES---KF---DNNNIYDNDDELEKKNLSDYISDVKNHVN 1065
 QY 928 EISNIKDTIFDNGVGLVKKNVNDAAHEVNTLNSAFFI-----QSLIEYNTTKESLNSLV 983
 DB 1066 NIYNIERGEDEEREFENKIQSTESHKSNE-----FICTENKSLRKQYMSKEDISNVI 1120
 QY 984 AMKVQVYLAQLSTGLNTIT-----DASKVELVSTALDETIDL--PTLSGLPLII 1032
 DB 1121 -----LKSDDINNLSKQNYPEILLDKQVMDNFQNNIEQNNDKLEDKLDG-AVF 1170
 QY 1033 ATIDGVSLGAAIKELSETNDPLLRQETAEKIGIMAVNLTAASTAIVTSALGIASGFSIL 1092
 DB 1171 EYLEDNKIIDSYKETNKENBELIKEYKLLKKNINIEIN----- 1208
 QY 1093 LVPLAGISAGIPSLVNNNELILODKATKVIDYF--KHI--SLABTEGAPTLDDK-----II 1144

```

Db 1209 -----DEMNDIKL-----LLNFFGIPYIQSPCEAEACSYLNNKNYCDALI 1250
Qy 1145 MPQDDVLASEIDFNNSITLGRKEIWAEGSGCHTLTDDIDHFFSSPSITYRKPWLSIYD 1204
Db 1251 SDDSDVLVF-----SGKTV---IKNFFN-----KKKTVEVYE 1279
Qy 1205 VLNIKKEKIDFSKDLMLPNAPNRVFGYEMGWTGPRSLDNDGTCLLDRIRDHYEGQFYW 1264
Db 1280 KKAIBEKGLYOELINI---SLGCGD--YTIGVHGIG----- 1313
Qy 1265 RYPAFTADAL-ITKLAPRYEDTNVRINLGNTRSFIVPVITTEQIRKNLSYSFYSGGSY 1323
Db 1314 -----IVNALEIIFKAPNFEDLKILKDIVSNPFRKIDKNYNEEQ-----FLNTHKNY 1363
Qy 1324 SLS-LSPYNN-----NIDLNVENDTWIDVNVKNI--TIESDEIQKGLI 1368
Db 1364 KLNWIFPNPPREVVKPKYKVCDDIKKFEWHVDPDIKSTFKLHKTNISE----- 1416
Qy 1369 ENILSKNIETDNKIIILNNHTINFYGDINESNRFISLTSILEDINIIIEIDLVSYSKYL 1428
Db 1417 EKVLNVN-----PILQYNN-----VRTYQSKIEDFPFELLE----- 1449
Qy 1429 LSGNCMKLIENSDDIOQKIDHIGNGEHOK 1458
Db 1450 -----KKRKTVDLIDHIRANNKQKR 1470

```

Search completed: November 5, 2003, 19:29:41
 Job time : 64.25 secs